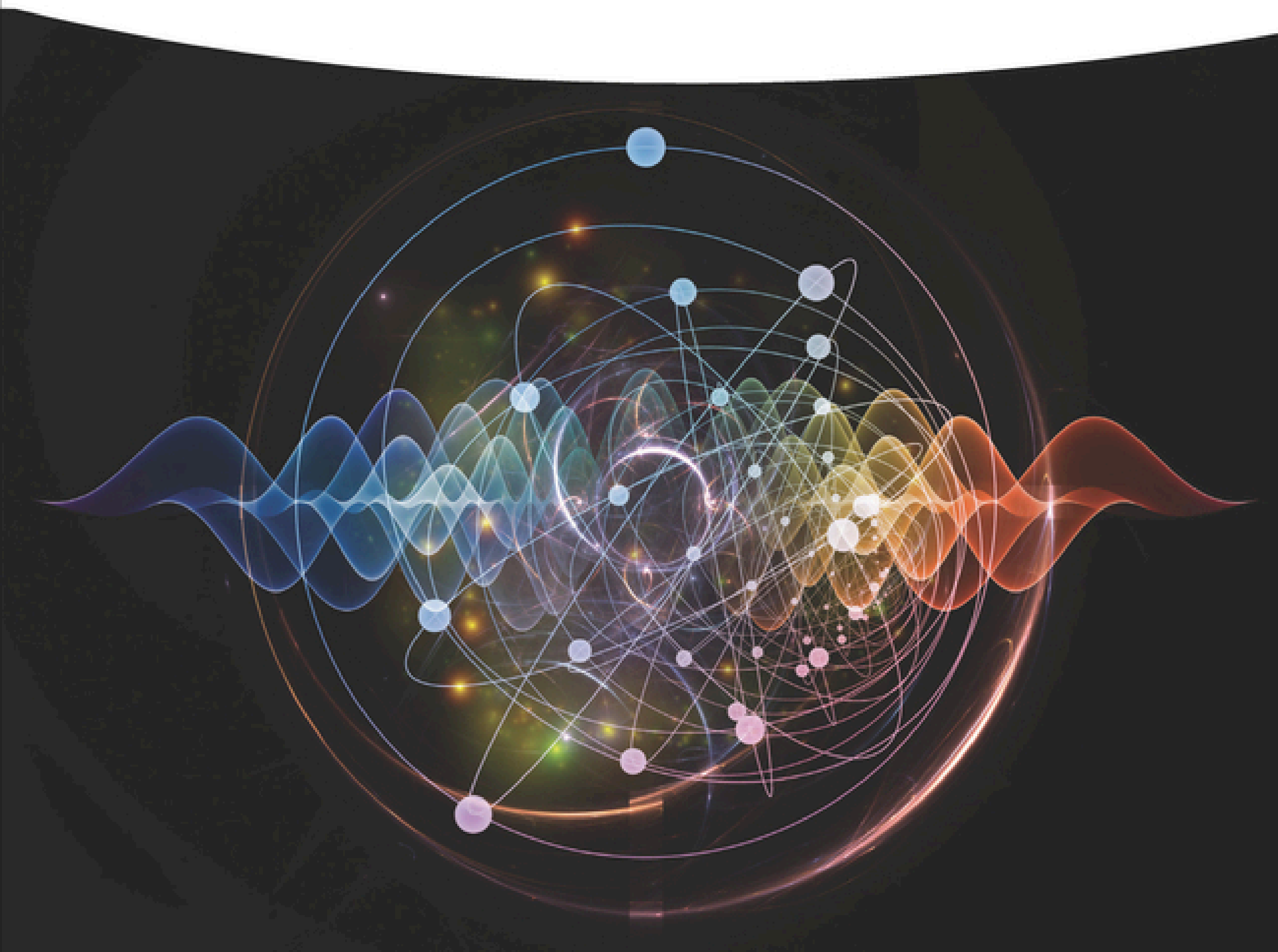
WILEY-VCH

James R. Chelikowsky

Introductory Quantum Mechanics with MATLAB

For Atoms, Molecules, Clusters and Nanocrystals



Introductory Quantum Mechanics with MATLAB®

Introductory Quantum Mechanics with MATLAB®

For Atoms, Molecules, Clusters, and Nanocrystals

James R. Chelikowsky

WILEY-VCH

Author
Prof. James R. Chelikowsky
University of Texas at Austin
Institute for Computational Engineering
and Sciences
Center for Computational Materials
201 East 24th Street
Austin, TX 78712
United States

Cover:

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

Print ISBN: 978-3-527-40926-6
ePDF ISBN: 978-3-527-65501-4
ePub ISBN: 978-3-527-65500-7
oBook ISBN: 978-3-527-65498-7

Cover Design Tata Consulting Services
Typesetting SPi Global, Chennai, India
Printing and Binding

Printed on acid-free paper

10 9 8 7 6 5 4 3 2 1

To Ellen

Contents

Preface

> *It is safe to say that nobody understands quantum mechanics.*
>
> Richard Feynman.
>
> *Those who are not shocked when they first come across quantum theory cannot possibly have understood it.*
>
> Niels Bohr.
>
> *The more success the quantum theory has, the sillier it looks.*
>
> Albert Einstein.

When some of the scientific giants of the twentieth century express doubt about a theory, e.g. referring to it as a "silly" or "shocking" theory, one might ask whether it makes sense to use such a theory in practical applications. Well, it does – because it works. Bohr, Einstein, and Feynman were reflecting on the exceptional nature of quantum theory and how it differs so strongly from previous theories. They did not question the success of the theory in describing the microscopic nature of matter. Indeed, quantum theory may be the most successful and revolutionary theory devised to date.

The theory of quantum mechanics provides a framework to accurately predict the spatial and energetic distributions of electronic states and the concurrent properties in atoms, molecules, clusters, nanostructures, and bulk liquids and solids. As a specific example, consider the electrical conductivity of elemental crystals. The ratio of the conductivities for a metal crystal such as silver to an insulating crystal such as sulfur can exceed 24 orders of magnitude. A comparable ratio is the size of our galaxy divided by the size of the head of a pin, i.e. an astronomical difference! Classical physics provides no explanation for such widely different conductivities; yet, quantum mechanics, specifically energy band theory, does.

That is the good news. The bad news is that quantum theory can be extraordinarily difficult to apply to real materials. There are a variety of reasons for this. One notable characteristic of quantum theory is the wave –particle duality of an electron, i.e. sometimes an electron behaves like a point particle, other times it behaves like a wave. When asked if an electron is a particle or a wave, an expert in quantum theory might respond with the seemingly incongruous answer: yes.

This is one reason why quantum theory seems so paradoxical and strange within the framework of traditional physics. Using quantum mechanics, we cannot ascertain the spatial characteristics of an electron by enumerating three spatial coordinates. Rather, quantum mechanics can only give us information on where the electron is likely to be, not where it actually is. The probabilistic nature of quantum theory adds many degrees of freedom to the problem. We need to specify a probability for finding an electron at every possible point in space. As a first pass, this issue alone would appear to doom any quantitative approach to solving for the quantum behavior of an electron.

Yet, the intellectual framework and the computational machinery have developed in spurts over the last several decades, which makes practical approaches to quantum mechanics feasible. Numerous textbooks often discuss the machinery of how to apply quantum mechanics, but fail to give the reader practical tools for accomplishing this goal. This is an odd situation whose origin likely resides in intellectual "latency." In the not too distant past, it would require a huge and complex computer code, run on a large mainframe, to do quantum mechanics for a relatively small molecule. This is no longer the case. A modest amount of computing power, such as that available with a laptop computer or in principle even a "smartphone," is sufficient to allow one to implement quantum theory for many systems of interest such as atoms, molecules, clusters, and other nanosacle structures. The goal of this book is to illustrate how this framework and machinery works. We will endeavor to give the reader a practical guide for applying quantum mechanics to interesting problems.

Many people are owed thanks to this effort. Yousef Saad helped me frame many of the electronic structure codes and gave guidance about algorithms for solving complex numerical problems. I also express a deep appreciation to a number of mentors and friends, including Marvin Cohen, Jim Phillips, and Steve Louie. Of course, I also thank my students and postdocs who did much of the heavy lifting.

Austin, Texas *James R. Chelikowsky*

1

Introduction

The great end of life is not knowledge but action.

– T. H. Huxley

1.1 Different Is Usually Controversial

Perhaps all breakthroughs in science are initially clouded with controversies. Consider the discovery of gravity. Isaac Newton invoked the concept of "action at a distance" when he developed his theory of gravity. Action at distance couples the motion of objects; yet, the objects possess no clear physical connection. Newton argued that the motion of an apple falling from a tree was similar to the motion of the moon falling toward (and fortunately missing) the earth. The source of the motion of both objects is consistent with an "action at a distance" caused by the presence of the earth and its gravitational field.

We can contrast the trajectory of the moon with a simpler object such as a golf ball. It is easy to understand that a golfer can make the ball move by striking it. A ball struck just right will carry hundreds of yards (or meters). Residents of Newton's time would be comfortable with this idea. The golf club directly contacts the ball, albeit for a very short time. The physical connection to the ball is the club swung by the golfer. But how can the earth change the moon's trajectory? The earth does not carry a big golf club to strike the moon. While the action at a distance theory may not be apparent to a lay person, or even a good scientist in Newton's time, the laws of gravity predicted the behavior of astronomical bodies such as the moon's orbit incredibly well. Hardly anyone would argue that we ignore the practical application of Newton's theory until someone resolved this action at distance business. For years, scientists argued the meaning of "action at distance" and the nature of space itself. Eventually, scientists agreed that the concept of Newtonian space was problematic. It was left to Einstein to straighten out issues of space, time, and gravity. In some sense, it hardly mattered if you wanted to predict planetary motion. A practical application of Newton's theory accomplished that really well, save some relatively minor fixes from Einstein. (We are not going to worry about issues such as worm holes or gravity waves.)

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

Perhaps one should view the theory of quantum mechanics in the same manner. The theory remains "mysterious" in some ways. Oddly, some of the central components of the theory are understandable only because we can think about them in classical terms. Still, quantum theory can be used to predict properties of matter with unprecedented accuracy. Upon the invention of quantum mechanics, the famous physicist Dirac wrote the following [1]:

> *The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems with out much computation …*

Dirac's quote of nearly a century ago is correct and appropriate. In principle, the underlying physical laws (of quantum mechanics) should allow one to predict "much of physics" (condensed matter physics as an example) and "the whole of chemistry." What stands in our way is a practical machinery to accomplish this task. Computational physicists often refer to this task as addressing *Dirac's Challenge*.

Dirac's warning about avoiding "too much computation" is an anachronism. Dirac's world of 1929 did not envision a computer capable of carrying out billions or trillions of mathematical operations within a second. When we contemplate the use of a modest computer, such as a laptop, for the work outlined in this book, it is indeed "without too much computation."

1.2 The Plan: Addressing Dirac's Challenge

Dirac's challenge for us is to develop "approximate practical methods of applying quantum mechanics." The goal of this book is to address, or better start to address, *the* challenge.

The book is roughly divided into three parts. The first part will focus on the theory. We will use a minimum of theory to get to the "good part." Our intent is not to write a quantum mechanics textbook. Rather, our intent in this part of the book is to review essential features. For example, we will consider the simplest of all atoms, hydrogen, and we will start with the simplest of theories from Bohr. We will then introduce the Schrödinger equation and briefly sketch out how to solve the hydrogen atom problem analytically.

The hydrogen atom is one of the rarest of quantum systems – one where we can do an analytical solution of the Schrödinger equation. The next several chapters will involve introducing the *real* problem, one with more than a single electron. A clear example of such a system is helium, where we have two electrons. Our study of the helium atom will lead us to consider the Hartree and Hartree–Fock approximations. Our next objective will be to consider a practical method for more than one or two electrons. A practical theory for this is based on "density functional theory," which focuses on how electrons are distributed in space.

A logical pathway to take us from Hartree–Fock theory to density functional theory arises from a "free electron" model. We introduce this model using concepts removed from the physics of an isolated atom. We will "backtrack" in our discussions to consider some solid-state physics concepts. Theories based on electron density will provide some key approximations. In particular, we will begin with the Thomas–Fermi approximation, which can lead to contemporary density functional theories. This approach will allow us to consider a "one-electron" Schrödinger equation to solve a many-electron problem.

The last chapter of this section will center on the "pseudopotential approximation." This key approximation will allow us to fix the length and energy scales of the many-electron problem by considering only the chemically relevant electronic states. The pseudopotential approximation treats an element such as lead on an equal footing with an element such as carbon. Both lead and carbon have the same configuration for the outermost, or valence, electrons. These chemically active states provide the chemical bond or "electronic glue" that holds atoms, clusters, molecules, and nanocrysals together.

The next part of the book illustrates numerical methods. Numerical methods are important as there are few atomic systems that can be solved *analytically*, save the aforementioned hydrogen atom. This is also true for classical systems where analytically only the two-body system is solvable.

We initially consider an isolated, spherically symmetric atom. We introduce the variational method and show how approximate wave functions can be used to obtain accurate estimates for the exact solution. We also solve the problem by integrating a one-dimensional equation.

We will consider solutions for many-electron atoms and molecules, using a numerical basis. This is the standard method for most quantum chemistry approaches to molecules and atoms, although it may not be the best method for these systems, especially for pedagogical purposes. An alternate is to solve the problem in real space on a grid. This approach is easy to implement and understand. With either a basis or a grid approach, we solve an "eigenvalue problem." Iterative methods can solve such problems and we will illustrate this.

The last part of the book demonstrates the application of quantum theory to atoms, molecules, and clusters using a common numerical method. Physical concepts such as pseudopotentials, density functional theory, and a real-space grid form the underpinnings for computing a solution of the electronic structure problem. The pseudopotential model of solids is widely used as the *standard model* for describing atomistic systems. The model divides electronic states into those that are valence states (chemically active) and those that are core states (chemically inert). For example, systems made up of silicon atoms have valence states derived from the atomic $3s^2 3p^2$ configuration. The valence states form bonds by promoting a 3s electron to form sp^3 bond. One can omit the core states $1s^2 2s^2 2p^6$ altogether in pseudopotential theory. As such, the energy and length scales for determining a basis are set by the valence state.

Density functional theory is perhaps the weakest approximation made in our toolbox, but it is indispensable. Density functional theory allows us to consider one electron at a time as it maps the difficult many-body problem to a one-electron problem.

Our use of a real-space grid reflects the bias of the author. The use of a grid to solve difficult differential equations is well known in the engineering community. There are good reasons for its popularity. Grids are easy to implement and possess a number of advantages for running on high-performance computers; e.g. they can reduce the number of global communications.

Many of the numerical solutions in our book are based on computer codes using MATLAB. The inventors of MATLAB claim: *MATLAB is a high-level language and interactive environment that enables you to perform computationally intensive tasks faster than with traditional programming languages such as C, C++, and Fortran*. There is merit to this claim; MATLAB is easy to use and implement. A transcription of a "state of the art" research program forms the basis of the codes in this book.

The application to atoms will focus on ionization energies, electron affinities, and polarizabilities. We will examine some diatomic molecules along with some organics such as methane and benzene. We will also examine solutions for chemical trends, ionicities, energy levels, bond energies, vibrational levels, and bond lengths. Clusters of atoms represent a "new" form of matter, i.e. a combination of atoms that is stable only in isolation. Systems in isolation represent a serious challenge for experiment. The systems must be probed without any material interactions. We can examine properties in clusters as for molecules, but we can also examine other seminal properties such as the evolution of properties from atoms to crystals. We will look at nanoscale structures such as nanocrystals. In these systems, quantum confinement can play an important role in altering physical properties. An appendix at the end will give the reader access to the essential codes.

Reference

1 Dirac, P.A.M. (1929). *Proceedings of the Royal Society of London Series A* 123: 714.

2

The Hydrogen Atom

"…to understand hydrogen is to understand all of physics…"

— V. Weisskopf

2.1 The Bohr Model

Before the days of quantum theory, workers focused on the relationship of light and matter. At the turn of the last century, Wilhelm Wein measured the intensity distribution of light as a function of frequency for blackbody radiation. He postulated a functional relationship for the intensity distribution; however, he provided no theoretical guidance for this function. Nonetheless, his work was so important that he won the 1911 Nobel Prize in physics.

The theoretical underpinning of Wien's work was later explained by Planck and Einstein. Planck postulated that the energy of light was quantized into discrete energy units, which he called "energy elements" and were later called "photons." (Actually much later – in the 1920s by Gilbert N. Lewis.) The energy of a photon is directly proportional to the frequency of the light:

$$E = Nhf = N\hbar\omega \tag{2.1}$$

where h is Planck's constant and f is the frequency of light ($\hbar = h/2\pi$ and ω is the angular frequency). N is the number of quanta.

As usual, Einstein's insights helped clarify the energy quanta of light. He wrote:

> *According to the assumption to be contemplated here, when a light ray is spreading from a point, the energy is not distributed continuously over ever-increasing spaces, but consists of a finite number of energy quanta that are localized in points in space, move without dividing, and can be absorbed or generated only as a whole.*

Some observers believe this statement by Einstein constitutes one of the most profound sentences written by a physicist in the last century [1, 2].

If light is quantized, what about matter? Are the properties of matter also quantized? Can we distinguish the two? Niels Bohr partly addressed this question

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

with a theory that accounted for the experimental spectrum of atomic hydrogen. Bohr's theory departed from the mainstream. Again, we remember that different can be controversial! Fortunately, Bohr was a superb scientist with impeccable credentials. A theory by a lesser scientist might receive ridicule instead of praise. Apparently this occurred. Arthur Haas in 1909 made a number of proposals in the field of atomic physics. One of his proposals anticipated Bohr's theory of the atom. However, prominent scientists of this era dismissed Haas' work and praised Bohr's theory [3, 4].

According to Bohr's theory of 1913, a hydrogen atom exists in a stationary state unless light is absorbed or emitted. When the state of the atom changes, the energy absorbed or emitted is quantized. We can reproduce Bohr's theory easily, albeit not exactly as he did. We will use some ideas developed some 10 years after Bohr's original work.

We will start from the assumption that electrons possess wave-like properties. Louis de Broglie developed this concept in 1922 [5]. He postulated that any particle moving has a wave length associated with its motion:

$$\lambda = \frac{h}{p} \tag{2.2}$$

where λ is the wave length, h is Planck's constant and p is the momentum of the particle. de Broglie's concept of a particle having a wave property is a central concept of quantum theory. Macroscopic bodies in motion possess large, very large, momenta. Their wavelengths are vanishingly small. We never associate a wave-like property with, say, a 1000-kg car moving at 100 km h^{-1}. The wavelength of such an object is $\sim 10^{-38}$ m, which is decidedly not measurable or observable. Particles at the microscopic scale are another story. Because their momenta can be so small, they can have a measurable wave length.

Bohr's theory postulated that an electron moves around the proton in circular orbits like a planet orbiting the sun. These orbits are not arbitrary and can take on only certain sizes. Suppose one assumes that such orbits exist when an integral number of de Broglie wave lengths equals the length of the orbit:

$$n\,\lambda = 2\pi r \tag{2.3}$$

where r is the orbital radius and n is an integer ($n = 1, 2, 3, \ldots$). Intuitively, this sounds right. It would be weird to have a characteristic wave length that was unrelated to a characteristic orbital length. Using de Broglie's wave length formula gives the allowed orbital radii:

$$r = n\frac{\hbar}{p} \tag{2.4}$$

The total energy of an electron in the hydrogen atom is the sum of the kinetic and potential energies:

$$E = \frac{p^2}{2m} - \frac{e^2}{r} \tag{2.5}$$

e is the charge of the electron, and m is the electron mass. The kinetic energy comes from the motion of the electron in orbit. The potential energy comes from the Coulomb energy of the positively charged proton interacting with the negatively charged electron.

We do not consider the kinetic energy of the proton as its mass is 1836 times that of the electron. As a first pass, we treat the proton's mass as infinite. This is a really good approximation. We rewrite this total energy expression by realizing that the radius, or the momentum (take your pick), is quantized.

$$E = \frac{n^2\hbar^2}{2mr^2} - \frac{e^2}{r} \tag{2.6}$$

The force, F, on the electron is given by the negative of the derivative of the total energy with position (as from elementary mechanics). The force vanishes when a particle is in equilibrium:

$$F = -\frac{\mathrm{d}E}{\mathrm{d}r} = -\frac{n^2\hbar^2}{mr^3} + \frac{e^2}{r^2} = 0 \tag{2.7}$$

This condition determines the stationary state radii:

$$r_n = \frac{n^2\hbar^2}{me^2} = 0.5292\, n^2 \text{ Å} \tag{2.8}$$

We can insert this value into Eq. (2.6) to get the allowed energies for an electron in a hydrogen atom:

$$E_n = -\frac{me^4}{2\hbar^2 n^2} = \frac{-13.606}{n^2} \text{ eV} \tag{2.9}$$

We find the allowed energy frequencies for light absorbed or omitted by a hydrogen atom using Planck's criterion. Suppose that the hydrogen atom is initially in state n_1, and after absorbing the energy of a photon is in state n_2. Energy conservation yields the following:

$$\Delta E = E_{n_2} - E_{n_1} = \hbar\omega_{n_1 n_2} = \frac{me^4}{2\hbar^2}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) \tag{2.10}$$

The spectral lines predicted by Eq. (2.10) are in perfect agreement with those observed experimentally and validated Bohr's theory. In Figure 2.1, we illustrate the differences between the hydrogen energy levels. The lines are labeled by the terminating level, e.g. the Lyman series all terminate in the lowest level ($n_1 = 1$), the Balmer series terminates in the second lowest level ($n_1 = 2$), and the Paschen series terminates in the third lowest level.

While the Bohr theory of the hydrogen atom dramatically systematized the atomic spectra and explained the interaction of light with the quantized atom states, the theory is woefully limited. First, it is only valid for atoms containing one electron. This includes ionized atoms (He^{+}, Li^{+2}, Be^{+3}, $\mathrm{B}^{+4}, \ldots$). Many-electron atoms, that is, atoms with more than one electron, starting with helium are not described by the Bohr theory. Second, the theory tells us about the energy levels, but fails to mention anything about how to compute the intensity of the spectral lines. As we shall see, the key to understanding this problem is symmetry. Some scientists, most notably Arnold Sommerfeld, attempted a generalization of Bohr's atomic model, e.g. considering elliptical instead of circular orbits, without much success.

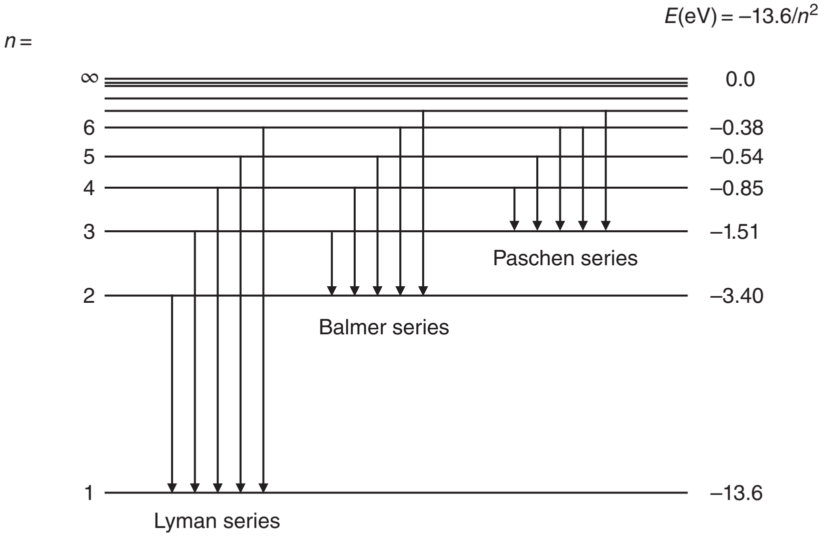


Figure 2.1 Atomic spectra of hydrogen atom illustrating the main transitions for three series. The Lyman transitions are in the ultraviolet region of the spectra. The Paschen transitions occur in the infrared. The levels are not to scale.

2.2 The Schrödinger Equation

Work by Erwin Schrödinger, Paul Dirac, and Werner Heisenberg ushered in the modern theory of quantum mechanics by replacing Bohr's theory with one capable of describing many-electron atoms. We will focus on Schrödinger's equation. Schrödinger's papers on "wave mechanics," written in 1926, resulted in quantum mechanical equations for predicting the properties of electrons interacting with nuclei [6]. A solution of his equation gives the correct energy levels for the hydrogen atom and much more. It yields an accurate description of many-electron atoms, and accounts for both the energetic and spatial distributions of electrons. Schrödinger's paper is one of the notable scientific achievements of the twentieth century.

For hydrogen, we can write the Schrödinger equation as

$$\left[\frac{-\hbar^2\nabla^2}{2m} - \frac{e^2}{r}\right]\Psi_n(\vec{r}) = E_n\Psi_n(\vec{r}) \tag{2.11}$$

The proton is placed at the origin. Again, we take the mass of the proton as infinite compared to the electron.

This Schrödinger equation constitutes an eigenvalue problem. We can label eigenvalues and eigenstates with "quantum numbers." We anticipate this by noting that n may stand for a set of such numbers. The eigenvalues, E_n, correspond to energy levels. The eigenstates provide us with a challenge. The physical interpretation of the eigenstates, Ψ_n, is one of *the* issues in quantum theory.

Some 80 years after Schrödinger wrote down this equation scientists still argue about what Ψ_n represents. One traditional, and widely held, view centers on a probabilistic interpretation. The eigenstates, or better the square of the eigenstates, give a measure for the probability of finding an electron at a particular point in space. Since the electron is somewhere, if we add up the probability of finding an electron in all space, we may write

$$P(\vec{r}) = |\Psi_n(\vec{r})|^2$$
$$\iiint P(\vec{r})\ \mathrm{d}^3 r = 1 \tag{2.12}$$

where P is the probability of finding an electron in a volume element $\mathrm{d}^3 r$. If we add up this probability by integrating over all space, it must add up to unity. We can always "normalize" the eigenstate so that the integral is one. This assumes that the eigenstate is integrable. Since our efforts are focused in practical issues, we are always going to assume that the eigenstates are well behaved, e.g. no discontinuities in the function or its derivatives, and no divergences or poles!

Eigenstates are often called "wave functions." The Schrödinger equation is then called a wave equation. We cannot measure wave functions as in the language of quantum theory they are not "observables." Observables are quantities that are measurable, e.g. the energy of an electron. Quantum theorists often say that we can kick an electron, but we cannot kick a wave function. The distinction between an observable and an non-observable object or property is not something a classical physicist would make out, although it is not a completely alien concept. Consider the concept of an imaginary number. We often use $i = \sqrt{-1}$ as a place holder for some mathematical operation that will lead to a "real number," but an intermediate operation might not. Roughly speaking, we are doing the same thing in quantum modeling. We might have a complex phase factor in our eigenstates that can take on different values. But when we square the eigenvalue the phase factor is not relevant; it is not an observable.

Einstein and others abhorred this situation with some variables being unobservables [7]. They also disliked a "probabilistic" interpretation of quantum mechanics. In a famous exchange, Einstein claimed: "God does not roll dice. The Lord is subtle, but not malicious." Bohr rejoined "Don't tell God what to do." (The quotes are probably not correct in detail, but they are right in spirit.) This is an odd set of circumstance. Here is a theory with great predictive power; yet, the interpretation of a central feature of the theory – the wave function – is opaque.

If we accept that the wave functions are related to probability, then we need to think about statistics and measurements. If someone claims that a particular event has a certain probability of happening, what does that mean? Traditionally, such a claim would entail that one must repeat the event over and over, and assess the various outcomes. For example, suppose we are given a coin and told that the coin is fair. What does that mean in practice? If I flip a coin once, I know nothing about whether the coin is fair. In contrast, if I flip the coin a thousand times and it comes up heads half the time, I am reasonably sure the coin is fair. If heads come up 90% of the time, I am very sure the coin is not fair. To assess this situation involves experiment (making a measurement) and recording the results (doing a statistical analysis).

For the case at hand, if one finds a quantum mechanical solution where the wave function localizes an electron within a given volume, this statement only makes sense if you do a series of measurements to determine where the electron is, e.g. by using X-rays. This relationship between measurement and the probabilistic interpretation of the wave functions lies at the heart of what quantum mechanics means.

This philosophical issue is vaguely similar to the question: Does a tree falling in the woods make a sound if no one is around to hear it? One might argue that if no one is around, the question is without meaning. The state of an electron is also without meaning if no one makes a measurement. We obtain information about reality only when we make a measurement or ask a question.

Heisenberg said: "What we observe is not nature itself, but nature exposed to our method of questioning [8]." If we make no measurement (or raise no question), we do not know about the sound of a falling tree, or the position of an electron. This subject is becoming too philosophical and we take leave of it. Our goal is more "practical."

Next we attempt to solve the Schrödinger equation for one of the simplest of systems. If we cannot solve this equation for a hydrogen atom, its interpretation is irrelevant. Many textbooks do the solution of a hydrogen atom in some detail, e.g. see Schiff [9]. Here we give a general outline of how it is done. A solution involves a number of variable transformations and the procedure can be slightly tedious.

The first step is to recognize that the potential energy for the electron–proton is spherical. Spherical coordinates are illustrated in Figure 2.2:

$$\begin{aligned} x &= r\ \cos\phi\ \sin\theta \\ y &= r\ \sin\phi\ \sin\theta \\ z &= r\ \cos\theta \end{aligned} \tag{2.13}$$

The wave function is separated into radial and angular components:

$$\Psi(r,\theta,\phi) = R(r)Y(\theta,\phi) \tag{2.14}$$

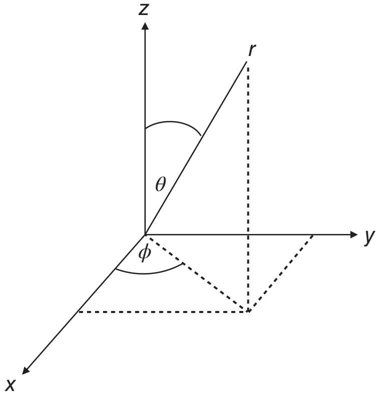


Figure 2.2 Spherical coordinates defined by (r, θ, ϕ).

When this wave function is inserted into Eq. (2.11) we get two equations. One depends on the radial coordinates, the other on the angular coordinates. This separation is possible because the potential is spherical. These equations are given by

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)+\frac{2m}{\hbar^2}\left(E+\frac{e^2}{r}\right)R-\frac{\lambda}{r^2}R=0 \tag{2.15}$$

and

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}+\lambda Y=0 \tag{2.16}$$

where λ is a constant introduced to separate the radial equation from the angular one. The equation for the angular part of the wave function is valid for *any* spherical potential. We can solve this equation in a straightforward manner following a prescription from nineteenth century mathematics [9]. The solution in spherical harmonics is:

$$Y_{lm}(\theta,\phi)=\sqrt{\frac{(2l+1)!}{4\pi}\frac{(l-m)!}{(l+m)!}}\,P_l^m(\cos\theta)\ \exp(im\phi) \tag{2.17}$$

The integers (l, m), may take on only certain values: $l = 0, 1, 2, \ldots$ and $m = -l, -l + 1, \ldots, -1, 0, 1, \ldots, l - 1, l$. The separation constant is given by $\lambda = l(l + 1)$. The P_l^m are called associated Legendre polynomials:

$$P_l^m(x)=\frac{(-1)^m}{2^l l!}(1-x^2)^{m/2}\frac{\mathrm{d}^{l+m}}{\mathrm{d}x^{l+m}}(x^2-1)^l \tag{2.18}$$

This expression is valid for $m \geq 0$. The following holds for $-m$:

$$P_l^{-m}(x)=(-1)^m\frac{(l-m)!}{(l+m)!}\,P_l^m(x)$$

Functions for the first few values of l are easy to list. We focus on angular functions with $l \leq 2$ as listed in Table 2.1.

The radial part of the wave function will depend on the nature of the potential. We can write

$$u(r)=rR(r) \tag{2.19}$$

We use this function in Eq. (2.15) to obtain

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2u}{\mathrm{d}r^2}+\frac{l(l+1)\hbar^2}{2mr^2}\,u-\frac{e^2}{r}\,u=E\,u \tag{2.20}$$

We define dimensionless variables, which are given by

$$\rho=\frac{\sqrt{8m|E|}}{\hbar}\,r \qquad \eta=\sqrt{\frac{m}{2|E|}}\,\frac{e^2}{\hbar} \tag{2.21}$$

The energy levels are bound in this analysis, so $E = -|E|$. We write Eq. (2.20) as

$$\frac{\mathrm{d}^2u}{\mathrm{d}\rho^2}-\frac{l(l+1)}{\rho^2}\,u+\left(\frac{\eta}{\rho}-\frac{1}{4}\right)\,u=0 \tag{2.22}$$

Table 2.1 Spherical harmonics.

l	m$_l$	Y_{lm}
0	0	$Y_{00} = \frac{1}{\sqrt{4\pi}}$
1	1	$Y_{11} = -\sqrt{\frac{3}{8\pi}} \sin\theta \exp\{(i\phi)\}$
1	0	$Y_{10} = \sqrt{\frac{3}{4\pi}} \cos\theta$
1	−1	$Y_{1,-1} = \sqrt{\frac{3}{8\pi}} \sin\theta \exp\{(-i\phi)\}$
2	2	$Y_{22} = \sqrt{\frac{15}{32\pi}} \sin^2\theta \exp\{(2i\phi)\}$
2	1	$Y_{21} = -\sqrt{\frac{15}{8\pi}} \cos\theta \sin\theta \exp\{(i\phi)\}$
2	0	$Y_{20} = \sqrt{\frac{5}{16\pi}} (3\cos^2\theta - 1)$
2	−1	$Y_{2,-1} = \sqrt{\frac{15}{8\pi}} \cos\theta \sin\theta \exp\{(-i\phi)\}$
2	−2	$Y_{2,-2} = \sqrt{\frac{15}{32\pi}} \sin^2\theta \exp\{(-2i\phi)\}$

This format allows us to establish the long range nature of the radial wave function. Consider $\rho \to \infty$.

$$\frac{d^2u}{d\rho^2} - \frac{1}{4} u = 0 \tag{2.23}$$

The function $u \to \exp(-\rho/2)$ in this limit. The sign of the exponential is obvious. Without a minus sign, we could not normalize the wave function and insure that the probability summed to unity. We build in the limiting case by writing

$$u = F(\rho) \exp(-\rho/2) \tag{2.24}$$

Upon insertion into Eq. (2.22), we obtain an equation for F:

$$\frac{d^2F}{d\rho^2} - \frac{dF}{d\rho} + \left[\frac{\eta}{\rho} - \frac{l(l+1)}{\rho^2}\right] F = 0 \tag{2.25}$$

We then express F as a polynomial as the asymptotic behavior is incorporated by $\exp(-\rho/2)$:

$$F(\rho) = \rho^s \sum_{k=0}^{\infty} a_k \rho^k = \rho^s (a_0 + a_1\rho + a_2\rho^2 + \cdots) \tag{2.26}$$

We will examine the long range and short range behavior of F to determine s and the coefficients a_k. We insert Eq. (2.26) into Eq. (2.25) and obtain:

$$\sum_{k=0}^{\infty} \left\{ [(k+s)(k+s-1) - l(l+1)]\rho^{k+s-2} - [(k+s) - \eta]\rho^{k+s-1} \right\} a_k = 0 \quad (2.27)$$

Consider $k = 0$ and the limit as $\rho \to 0$:

$$\{[s(s-1) - l(l+1)]\}a_0 = 0 \quad (2.28)$$

For a non-zero value of a_0, $s(s-1) = l(l+1)$. A solution yields $s = l+1$ or $s = -l$. If R is finite as $\rho \to 0$ (or $r \to 0$), we need $s = l+1$.

We can now establish a relationship that will dictate the long range behavior of F by considering equal powers of ρ:

$$\sum_{k=0}^{\infty} \left\{ [(k+l+1)(k+l) - l(l+1)]\rho^{k+l-1} - [(k+l+1) - \eta]\rho^{k+l} \right\} a_k = 0$$

$$\sum_{k=0}^{\infty} \left\{ [(k+l+1)(k+l) - l(l+1)]a_k - [(k+l) - \eta]a_{k-1} \right\} \rho^{k+l-1} = 0 \quad (2.29)$$

If this is to hold for arbitrary values of ρ, we have a relationship between a_k and a_{k-1}:

$$\frac{a_k}{a_{k-1}} = \frac{(k+l) - \eta}{[(k+l+1)(k+l) - l(l+1)]} \quad (2.30)$$

This expression has the following behavior for large k:

$$\lim_{k\to\infty} \frac{a_k}{a_{k-1}} = \lim_{k\to\infty} \frac{(k+l) - \eta}{[(k+l+1)(k+l) - l(l+1)]} \to \frac{1}{k} \quad (2.31)$$

We consider an expansion for $\exp(\rho)$ in powers of ρ^k:

$$\exp(\rho) = \sum_{k=0}^{\infty} \frac{1}{k!}\rho^k = \sum_{k=0}^{\infty} a_k \rho^k \quad (2.32)$$

This gives values of a_k for large k:

$$\lim_{k\to\infty} \frac{a_k}{a_{k-1}} = \lim_{k\to\infty} \frac{(k-1)!}{k!} = \frac{1}{k} \quad (2.33)$$

This behavior for $\exp(\rho)$ is crucial to rendering a solution. The polynomial expansion for F for large powers of k are the same as $\exp(\rho)$ and since $u = F\ \exp(-\rho/2)$ u diverges at large ρ (or r). We cannot allow that and still be able to normalize the wave function. The wave function must go to zero as $\rho \to \infty$. One way to resolve this "bad behavior" is to make sure the series terminates. If $F \sim \rho^m$ where m is a finite positive integer, then $u = F(\rho)\exp(-\rho/2) \to 0$ as $\rho \to \infty$ and we are good.

The polynomial series for F (Eq. (2.30)) will terminate when

$$\eta = k + l = n; \quad n = 1, 2, \ldots \quad (2.34)$$

Recalling the definition of η yields the following:

$$E_n = -\frac{me^4}{2n^2\hbar^2} \quad (2.35)$$

The energy levels are the same as given by the Bohr model. (Bohr had the better of it as he never even had to deal with a complex differential equation.)

Quantum numbers are used to label each electronic state. In our solution for the hydrogen atom there are three integers called spatial quantum numbers: *nlm*. One quantum number exists for each spatial dimension, i.e. three dimensional-space yields three quantum numbers. The energy levels of the electron in a hydrogen atom depend only on *n*, the *principal quantum number* because the potential is spherically symmetric or independent of the angular coordinates. Besides using *n*, we label the electronic states by the *angular* quantum number, *l*, and the *azimuthal* quantum number, *m*.

The lowest energy state occurs for $n = 1$ with $l = m = 0$. The radial part of the solution is given by our expression for u and F (see Eqs. (2.19), (2.24), and (2.26)):

$$R_{10}(r) = \frac{2}{\sqrt{a_0^3}} \exp(-r/a_0) \tag{2.36}$$

where a_0 is the Bohr unit (see the Appendix A and Eq. (2.8) where $a_0 = r_1$).

These radial functions involve *associated Laguerre polynomials*. We summarize the radial functions for the first few values of the hydrogen quantum numbers in Table 2.2. We generalized the functions for a “hydrogen-like” atom. We replace e^2 by Ze^2, where Z is the atomic number. This generalization allows us to consider any “one-electron” atom using the Bohr model.

We can compose the total wave functions for any set of quantum numbers, *nlm*, using Tables 2.1 and 2.2. The results are normalized so that the integrals over the radial part and angular are both unity.

$$\iiint |\psi_{nlm}(\vec{r})|^2 \, d^3r = \int |R_{nl}(\vec{r})|^2 \, r^2 \, dr \iint |Y_{lm}(\theta, \phi)|^2 d\Omega = 1 \tag{2.37}$$

We depict the hydrogen wave functions (also known as orbitals) by constructing isosurface plots as given by Figure 2.3. We show the first allowed state for each value of *l*. The spatial configuration of these orbitals when combined with other orbitals is related to the nature of the chemical bond.

Table 2.2 Radial part of the wave functions for hydrogen-like atoms with Z protons in the nucleus.

n	*l*	R_{nl}
1	0	$2\sqrt{(Z/a_0)^3}\ \exp(-Zr/a_0)$
2	0	$2\sqrt{(Z/2a_0)^3}\ (1 - Zr/2a_0)\ \exp(-Zr/2a_0)$
2	1	$\sqrt{(Z/2a_0)^3}\ (Zr/\sqrt{3}a_0)\ \exp(-Zr/2a_0)$
3	0	$2\sqrt{(Z/3a_0)^3}\ [1 - 2Zr/3a_0 + 2(Zr/a_0)^2/27]\ \exp(-Zr/3a_0)$
3	1	$\sqrt{(Z/3a_0)^3}\ (4\sqrt{2}Zr/3a_0)\ [1 - Zr/6a_0]\ \exp(-Zr/3a_0)$
3	2	$(2/27)\ \sqrt{2/5}\ \sqrt{(Z/3a_0)^3}\ (Zr/a_0)^2\ \exp(-Zr/3a_0)$

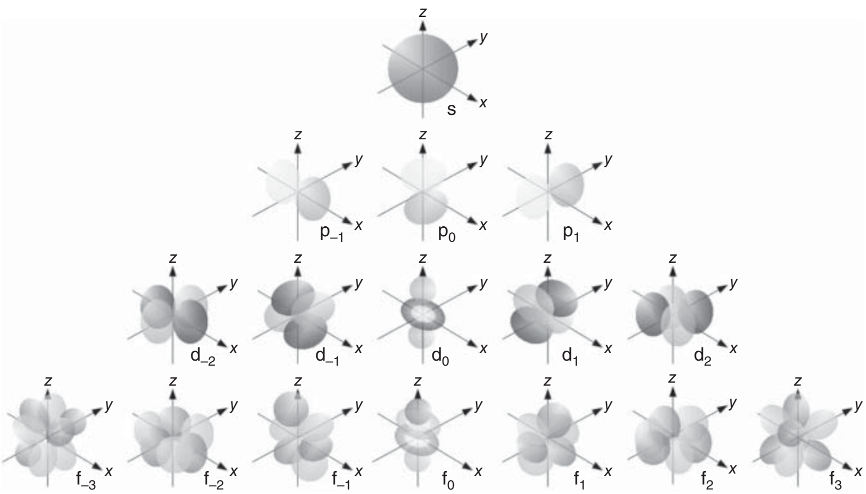


Figure 2.3 Spatial extent of the atomic wave functions for a hydrogen atom.

2.3 The Electronic Structure of Atoms and the Periodic Table

The quantum description of the hydrogen atom offers an explanation not for the structure of atoms, but more – the periodic table. One can construct the periodic table by considering the following:

The Pauli exclusion principle. The Pauli exclusion principle states that one electron and one electron alone can have the same set of quantum numbers.

The independent electron approximation. This approximation assumes that each electron can be discussed individually, e.g. each electron moves in an average field of the other electrons. To the extent this approximation holds, we can ascribe hydrogen atom characteristics to many-electron atoms. That we can make such an assumption says something about electrons interacting among themselves. We will shortly capitalize on this assumption.

The spin of the electron. Another quantum number is lurking out there – one that has no analogy to the spatial quantum numbers. Given one quantum number per spatial degree of freedom, we cannot associate another number with three-dimensional space. However, nature says otherwise. We can assign two electrons with the same spatial quantum numbers (n, l, m). This violates the Pauli exclusion principle, but not if we had another quantum number! Let us assume we do. We assign this additional quantum number to electron "spin." We are not capricious in adding another quantum number, although maybe nature is. We need to do so or we cannot organize electronic states.

We associate spin with a magnetic moment of the electron. The spin quantum number of the electron was first detected by Gerlach and Stern using a strong, inhomogeneous magnetic field [10]. They deflected a beam of electrons with such a field and found the beam deflected up or down in equal proportions. They demonstrated that quantum angular momentum of any kind has a discrete spectrum. One could say they discovered that angular momentum is quantized. We interpret their measurement as a demonstration of electron spin. The quantized value of this spin, $\vec{S}$, is $\pm\hbar/2$. The magnetic moment associated with this spin is

$$\vec{\mu}_{\rm s} = -\frac{g_{\rm s}\mu_{\rm B}}{\hbar}\vec{S}$$

where $g_{\rm s}$ is the electron spin constant, $g_{\rm s} \approx 2.0023$, $\mu_{\rm B}$ is the Bohr magneton, $e\hbar/2mc$, m is the electron, c is the speed of light, and e is the charge on the electron. This expression is in Gaussian CGS units (see the appendix for a discussion of units.)

Does the magnetic moment of the electron arise from a "spinning electron," e.g. a charged spherical electron spinning on its axis? The answer is a negative. Suppose we assign the size of the electron by $r_{\rm e} = e^2/mc^2 = 2.818 \times 10^{-13}$ cm, which is termed the "classical electron radius," and is also known as the "Lorentz radius."

Assume that an object of this size is spinning about some axis. While the idea of the "self-rotation of the electron" naturally gives rise to an angular momentum vector in addition to the one associated with orbital motion (quantum numbers l, m), the electron, owing to its small size, must spin really, really fast. Really, really fast implies its velocity at the equator (assuming a spherical electron) must exceed the speed of light in order to generate the observed magnetic moment. This violates the cardinal rule that nothing moves faster than light, or better, nothing transmits information faster than the speed of light. We will not consider whether the rotation of an electron can transfer data. In any event, quantum theory does not allow such a naive interpretation. For our purposes, we follow the field and accept the existence of an additional quantum number, which is called – *spin*.

The labels used for quantum states in the hydrogen atom now include the principal (n), orbital (l), azimuthal (m), and spin ($m_{\rm s}$) quantum numbers. The principal quantum numbers take on values of $n = 1, 2, 3, \ldots$. The orbital quantum numbers take on values of $l = 0, 1, 2, \ldots, n-1$. Orbital notation using letters $(s, p, d, f, g, h, \ldots)$ are often used instead of the numbers. The letter notation comes from the associated spectral features called sharp (s or $l = 0$), principal (p or $l = 1$), diffuse (d or $l = 2$), fundamental (f or $l = 2$). For $l \geq 3$, the letters are in alphabetical order, e.g. $l = 3, 4, 5, \ldots$ corresponds to $g, h, i, \ldots$.

We list the electronic configurations for atoms up to atomic number 56 in Table 2.3. We base the configurations on the quantum numbers associated with the hydrogen atom and fill up the levels with the Pauli principle.

The resulting configurations are often appropriate for atoms with more than one electron. The configurations are given by the principal quantum number, the orbital quantum number, and the occupancy of the state, e.g. we characterize Li atom by two electrons with $n = 1, l = 0$ and one electron with $n = 2, l = 0$ or $1s^2 2s^1$. As the number of electrons increase, we first consider the $n = 1$ shell, i.e. filling all the states with this principal quantum number; then we fill the $n = 2$ shell and continue the process. We consider the shells up to $n = 5$ and start the first two elements for $n = 6$.

Table 2.3 Electronic configurations for atoms up to atomic number 56.

Atomic number	Element	Configuration	Atomic number	Element	Configuration
1	Hydrogen	$1s^1$	30	Zinc	[Ar + $3d^{10}$]$4s^2$
2	Helium	$1s^2$	31	Gallium	[Ar + $3d^{10}$]$4s^2 4p^1$
3	Lithium	[He] $2s^1$	32	Germanium	[Ar + $3d^{10}$]$4s^2 4p^2$
4	Beryllium	[He] $2s^2$	33	Arsenic	[Ar + $3d^{10}$]$4s^2 4p^3$
5	Boron	[He] $2s^2 2p^1$	34	Selenium	[Ar + $3d^{10}$]$4s^2 4p^4$
6	Carbon	[He] $2s^2 2p^2$	35	Bromine	[Ar + $3d^{10}$]$4s^2 4p^5$
7	Nitrogen	[He] $2s^2 2p^3$	36	Krypton	[Ar + $3d^{10}$]$4s^2 4p^6$
8	Oxygen	[He] $2s^2 2p^4$	37	Rubidium	[Kr]$5s^1$
9	Fluorine	[He] $2s^2 2p^5$	38	Strontium	[Kr]$5s^2$
10	Neon	[He] $2s^2 2p^6$	39	Yttrium	[Kr]$5s^2 4d^1$
11	Sodium	[Ne] $3s^1$	40	Zirconium	[Kr]$5s^2 4d^2$
12	Magnesium	[Ne] $3s^2$	41	Niobium	[Kr]$5s^1 4d^4$
13	Aluminum	[Ne] $3s^2 3p^1$	42	Molybdenum	[Kr]$5s^1 4d^5$
14	Silicon	[Ne] $3s^2 3p^2$	43	Technetium	[Kr]$5s^1 4d^6$
15	Phosphorous	[Ne] $3s^2 3p^3$	44	Ruthenium	[Kr]$5s^1 4d^7$
16	Sulfur	[Ne] $3s^2 3p^4$	45	Rhodium	[Kr]$5s^1 4d^8$
17	Chlorine	[Ne] $3s^2 3p^5$	46	Palladium	[Kr]$4d^{10}$
18	Argon	[Ne] $3s^2 3p^6$	47	Silver	[Kr]$5s^1 4d^{10}$
19	Potassium	[Ar] $4s^1$	48	Cadmium	[Kr + $4d^{10}$]$5s^2$
20	Calcium	[Ar] $4s^2$	49	Indium	[Kr + $4d^{10}$]$5s^2 5p^1$
21	Scandium	[Ar] $4s^2 3d^1$	50	Tin	[Kr + $4d^{10}$]$5s^2 5p^2$
22	Titanium	[Ar] $4s^2 d^2$	51	Antimony	[Kr + $4d^{10}$]$5s^2 5p^3$
23	Vanadium	[Ar] $4s^2 3d^3$	52	Tellurium	[Kr + $4d^{10}$]$5s^2 5p^4$
24	Chromium	[Ar] $4s^1 3d^5$	53	Iodine	[Kr + $4d^{10}$]$5s^2 5p^5$
25	Manganese	[Ar] $4s^2 4d^5$	54	Xenon	[Kr + $4d^{10}$]$5s^2 5p^6$
26	Iron	[Ar] $4s^2 3d^6$	55	Cesium	[Xe] $6s^1$
27	Cobalt	[Ar] $4s^2 3d^7$	56	Barium	[Xe] $6s^2$
28	Nickel	[Ar] $4s^2 3d^8$			
29	Copper	[Ar] $4s^1 3d^{10}$			

The inner shells are indicated by the corresponding inert gas configuration.

The *periodic table of the elements* is based on the filling of these hydrogen-like states. The elements are organized in groups based on the outermost states; e.g. carbon, silicon, germanium, and tin all possess four outer electronic states: s^2p^2 and have similar elemental structures and chemical properties.

We are almost all right, but life or science is rarely this easy. Some anomalies exist in Table 2.3 that reflect the deficiencies of applying hydrogenic states to many-electron atoms. One of the most notable anomalies occurs for the potassium atom. One might guess that potassium has the electronic configuration of the Ar atom and one outer electron in the d orbital: [Ar] $3d^1$. That is what a hydrogen atom would do. Instead, nature gives potassium a configuration of [Ar] $4s^1$, i.e., we find the s-shell fills *before* the 3d shell.

Other anomalies occur in the filling of the 3d states. The occupancy does not begin with $3d^1$ and end with $3d^{10}$ by increasing the number of electrons in the *d* states linearly with the increasing atomic number, e.g. we have vanadium with a configuration of $4s^2 3d^3$ and chromium with $4s^1 3d^5$. A configuration of $4s^2 3d^4$ never happens! This absence of the $4s^2 3d^4$ configuration occurs because of the special stability of half-filled shells. We will address some of these anomalies in the next chapter when we consider many-electron atoms.

References

1 Bethe, H.A. and Salpeter, E.E. (1957). *Quantum Mechanics of One- and Two-Electron Atoms*. Springer-Verlag.

2 Folsing, A. (1994, 1997). *Albert Einstein: A Biography*. New York: Penguin Press Science.

3 Hermann, A. (1974). *The Genesis of Quantum Theory (1899–1913)*, MIT Press.

4 Bohr, N. (1913). On the constitution of atoms and molecules. *Philosophical Magazine* 26: 1, 475, and 857.

5 de Broglie, L. (1924). Recherches sur la théorie des quanta (Researches on the quantum theory). thesis. University of Paris, France.

6 Schrödinger, E. (1926). An undulatory theory of the mechanics of atoms and molecules. *Physical Review* 28: 1049.

7 Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review* 47: 777.

8 Heisenberg, W. (1999). *Physics and Philosophy: The Revolution in Modern Science*. Prometheus Books.

9 Schiff, L. (1968). *Quantum Mechanics*. 3e. McGraw-Hill.

10 Gerlach, W. and Stern, O. (1922). Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld. *Zeitschrift fur Physik* 9, 349.

3

Many-electron Atoms

> *Fools ignore complexity. Pragmatists suffer it. Some can avoid it. Geniuses remove it.*
>
> — Alan J. Perlis

3.1 The Variational Principle

We need to adopt a strategy that will give us hope for solving one of life's most difficult problems – obtaining the electronic structure of a many-electron atom! We establish a procedure for evaluating an approximate solution to the Schrödinger equation.

Suppose we know the solution for an atom using the Schrödinger equation:

$$\mathcal{H}\Psi_n = E_n\Psi_n \tag{3.1}$$

where the wave function corresponds to some quantum state and the energy corresponds to bound energy levels. Let us call the lowest energy state, E_0. Notation issue: we write $E_0 = -|E_0|$. We expect the energy of the state to be less than zero, which corresponds to a "bound state." If we know the wave functions corresponding to this state, we can write

$$E_0 = \int \Psi_n^* \mathcal{H} \Psi_n \mathrm{d}^3 r \tag{3.2}$$

This expression could correspond to the solution of a hydrogen atom where the wave function corresponds to a 1s state, and $E_0 = -1$ Ry $= -13.6$ eV. We will assume that we have but one electron in our initial discussion.

Suppose we have an approximate wave function, Ψ, which is a guess at the lowest energy level. We assume this approximate state is normalized, i.e. the state is a run of the mill function with nothing pathological about it – no poles, no discontinuities, integrable, … what have you. We write

$$\hat{E} = \int \Psi^* \mathcal{H} \Psi \mathrm{d}^3 r \tag{3.3}$$

The *variational principle* states that $\hat{E} \geq E_0$, i.e. the approximate energy level is higher in energy (less negative) than the true energy level.

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

We can confirm this. We write the approximate wave function as

$$\Psi = \sum_n a_n \Psi_n \tag{3.4}$$

This is a valid expression that holds for any approximate function, provided the approximate function is well behaved as we have assumed. We plug this expansion into Eq. (3.2) and obtain

$$\hat{E} = \int \sum_n a_n^* \Psi_n^* \mathscr{H} \sum_m a_m \Psi_m \, \mathrm{d}^3 r = \sum_{nm} a_n^* a_m E_m \int \Psi_n^* \Psi_m \, \mathrm{d}^3 r \tag{3.5}$$

We permit the wave function to be complex and consider the complex conjugates to find absolute values.

We now assume an orthogonality requirement for the wave functions:

$$\int \Psi_n^* \Psi_m \, \mathrm{d}^3 r = \delta_{nm} \tag{3.6}$$

where $\delta_{nm} = 1$ if $m = n$ and is otherwise zero. Should the wave functions not be orthogonal, we can always make them so. This allows us to write

$$\hat{E} = \sum_n |a_n|^2 E_n = |a_0|^2 \, E_0 + |a_1|^2 \, E_1 + |a_2|^2 \, E_2 + \cdots \tag{3.7}$$

We can rewrite this as

$$\hat{E}/E_0 = |a_0|^2 + |a_1|^2 \, E_1/E_0 + |a_2|^2 \, E_2/E_0 + \cdots \tag{3.8}$$

Now recall that the E_0 eigenvalue is the largest one. As such, $E_n/E_0 < 1$.

If we normalize the approximate wave function, we can write

$$1 = |a_0|^2 + |a_1|^2 + |a_2|^2 + \cdots \tag{3.9}$$

and this implies

$$1 \geq |a_0|^2 + |a_1|^2 \, E_1/E_0 + |a_2|^2 \, E_2/E_0 + \cdots = \hat{E}/E_0 \tag{3.10}$$

as the E_n/E_0 factor reduces the size of each $|a_n|^2$ term in the sum. Remembering the sign convention ($E_0 = -|E_0|$ and $\hat{E} = -|\hat{E}|$)

$$\hat{E} \geq E_0 \tag{3.11}$$

QED! Our approximate wave function will have a higher energy (again, a less negative energy) than the true one. As we improve the quality of our approximate wave function, we will approach the true energy from above.

This is only a sketch of the variational principle as the full theory is more general than we have indicated and holds for wave functions describing more than one electron. We will test the outline of our derivation to get some confidence in how well it works.

3.1.1 Estimating the Energy of a Helium Atom

What is the simplest atom with more than one electron? Helium. Thus, we consider a helium atom. We write down the Schrödinger equation for this atom as follows:

$$\mathscr{H}\Psi = \left[\frac{-\hbar^2\nabla_1^2}{2m} + \frac{-\hbar^2\nabla_2^2}{2m} - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{|\vec{r}_1 - \vec{r}_2|}\right]\Psi = E\Psi \tag{3.12}$$

We illustrate the coordinate system for the helium atom in Figure 3.1. This atom is more complex than hydrogen because it involves the coordinates of the two electrons $(\vec{r}_1, \vec{r}_2)$ and the Coulomb repulsion between the electrons. This latter term is annoying as it couples the two coordinates. If we were to neglect this term, we can write

$$\mathscr{H}\Psi = \left[\frac{-\hbar^2\nabla_1^2}{2m} + \frac{-\hbar^2\nabla_2^2}{2m} - \frac{2e^2}{r_1} - \frac{2e^2}{r_2}\right]\Psi = E\Psi \tag{3.13}$$

This equation is easy to solve. We can take the wave function as $\Psi(\vec{r}_1, \vec{r}_2) = \psi(\vec{r}_1)\psi(\vec{r}_2)$ and solve one equation:

$$\mathscr{H}\psi = \left[\frac{-\hbar^2\nabla^2}{2m} - \frac{2e^2}{r}\right]\psi = E\psi \tag{3.14}$$

This is nothing more than our hydrogen atom with the nuclear charge changed from one proton to two protons. We can immediately write down the energy of this system (remembering that we have two electrons):

$$E = -\frac{m(2e^2)^2}{2\hbar^2} - \frac{m(2e^2)^2}{2\hbar^2} = -\frac{4me^2}{\hbar^2} \tag{3.15}$$

The two noninteracting electrons have an energy of −8 Ry or −108.8 eV.

We can establish what this energy should be from experiment. The ionization energy, I, for a helium atom is measured to be 24.5 eV. This energy corresponds to the difference between a neutral helium atom and an ionized atom: $I = E(\text{He}^+) - E(\text{He})$. We know the precise energy of an ionized helium atom from the Bohr model or from the Schrödinger equation. The energy is that of a hydrogen atom with a nuclear charge changed to two protons: $E(\text{He}^+) = -54.4$ eV. As such, we have $E(\text{He}) = E(\text{He}^+) - I = -54.4 - 24.5 = -78.9$ eV. Our estimate of −108.8 eV neglects electron–electron repulsion, so it fails to yield a reliable value, giving an error of over 20%.

We can take a guess at what the interaction energy might be from a Bohr-like atom. Suppose the electrons are opposite one another as they orbit the nucleus.

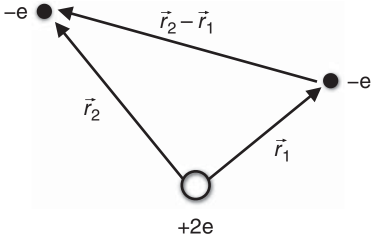


Figure 3.1 Atom coordinates for the helium atom. Two protons exist in the nucleus with the net charge being +2e.

This geometry would reduce the repulsive Coulomb energy between the electrons, but maintain the attractive Coulomb energy between each electron and the nucleus. If the electron separation maintains the Bohr radius given by $2a_0$, the interaction energy is $1/2a_0 + 1/2a_0 = 1/a_0$ or one Hartree (in atomic units – see appendix for more on units). One Hartree is 27.2 eV and the total energy would be -108.8 eV+ 27.2 eV= -81.6 eV versus the experimental value of -78.9 eV. Not a bad guess! The error is only ~3% and suggests we are on the right track.

We attempt to do better by estimating the repulsive energy with the variational principle. Let us write an approximate wave function using the results from a hydrogen-like atom from Tables 2.1 and 2.2:

$$\begin{aligned}\Psi(\vec{r}_1,\vec{r}_2) &= \psi_{10}(r_1)\psi_{10}(r_2)\\ &= \frac{1}{\sqrt{\pi}}\sqrt{(2/a_0)^3}\ \exp(-2r_1/a_0)\frac{1}{\sqrt{\pi}}\sqrt{(2/a_0)^3}\ \exp(-2r_2/a_0)\end{aligned} \tag{3.16}$$

where a_0 is the Bohr unit, $a_0 = \hbar^2/me^2 = 0.5292$ Å. These wave functions yield an estimate for the energy:

$$\begin{aligned}\hat{E} = &\iint (8/\pi a_0^3)\ \exp(-2(r_1+r_2)/a_0)\\ &\times\left[-\frac{\hbar^2\nabla_1^2}{2m} - \frac{-\hbar^2\nabla_2^2}{2m} - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{|\vec{r}_1-\vec{r}_2|}\right]\\ &\times(8/\pi a_0^3)\ \exp(-2(r_1+r_2)/a_0)d^3r_1d^3r_2\end{aligned} \tag{3.17}$$

With the notable exception of that annoying Coulomb repulsion term, the integrals are straightforward and yield the energy corresponding to hydrogenic atoms as before. We can evaluate the repulsive term:

$$(8e/\pi a_0^3)^2\iint \frac{\exp(-4(r_1+r_2)/a_0)}{|\vec{r}_1-\vec{r}_2|}\,d^3r_1d^3r_2 = \frac{5e^2}{4a_0} = 34.0\text{ eV} \tag{3.18}$$

Our estimate for the total electronic energy of a helium atom is now $E(\text{He}) < \hat{E} = -108.8 + 34.0 = -74.8$ eV. This estimate is too high by almost 4 eV or a ~5% error. This is no better than our crude guess by maximizing the electron–electron distance. At least it is a more rigorous estimate, and it is much better than ignoring the electrostatic repulsion altogether. A better approximation for the wave functions will give us a more accurate value, as illustrated in Chapter 7.

3.2 The Hartree Approximation

While the He atom is of some import, it only contains two electrons. We need to consider more complex systems to make progress for many-electron atoms. Consider a general problem of N nucleons of charge Z_n at positions $\{\vec{R}_n\}$ for $n = 1, \ldots,$ N and M electrons at positions $\{\vec{r}_i\}$ for $i = 1, \ldots, M$. This is shown schematically in Figure 3.2. The Hamiltonian for this system in its simplest

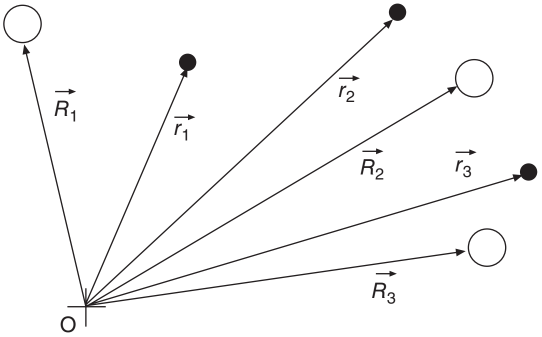


Figure 3.2 Atomic and electronic coordinates. The electrons are illustrated by filled circles and the nuclei by open circles.

form is

$$\mathscr{H}\left(\vec{R}_1,\vec{R}_2,\vec{R}_3,\ldots;\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots\right)=\sum_{n=1}^{N}\frac{-\hbar^2\nabla_n^2}{2\mathscr{M}_n}+\frac{1}{2}\sum_{n,m=1,n\neq m}^{N}\frac{Z_nZ_m\mathrm{e}^2}{|\vec{R}_n-\vec{R}_m|}$$
$$+\sum_{i=1}^{M}\frac{-\hbar^2\nabla_i^2}{2m}-\sum_{n=1}^{N}\sum_{i=1}^{M}\frac{Z_n\mathrm{e}^2}{|\vec{R}_n-\vec{r}_i|}+\frac{1}{2}\sum_{i,j=1,i\neq j}^{M}\frac{\mathrm{e}^2}{|\vec{r}_i-\vec{r}_j|} \qquad (3.19)$$

where $\mathscr{M}_n$ is the mass of the nucleon, which we often taken to be infinite as we did for the hydrogen and helium atoms. This expression omits some terms such as those involving relativistic interactions, but captures the essential features. Using the Hamiltonian in Eq. (3.19), we write the quantum mechanical equation for the electronic structure of the system as follows:

$$\mathscr{H}\left(\vec{R}_1,\vec{R}_2,\vec{R}_3,\ldots;\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots\right)\Psi\left(\vec{R}_1,\vec{R}_2,\vec{R}_3,\ldots;\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots\right)$$
$$=E\Psi\left(\vec{R}_1,\vec{R}_2,\vec{R}_3,\ldots;\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots\right) \qquad (3.20)$$

We estimate the total energy using a generalization of the variational theorem, provided we have a reliable wave function:

$$E=\frac{\int\Psi^*\mathscr{H}\Psi\,\mathrm{d}^3R_1\,\mathrm{d}^3R_2\,\mathrm{d}^3R_3\ldots\,\mathrm{d}^3r_1\,\mathrm{d}^3r_2\,\mathrm{d}^3r_3\ldots}{\int\Psi^*\Psi\,\mathrm{d}^3R_1\,\mathrm{d}^3R_2\,\mathrm{d}^3R_3\ldots\,\mathrm{d}^3r_1\,\mathrm{d}^3r_2\,\mathrm{d}^3r_3\ldots} \qquad (3.21)$$

Solving Eq. (3.20) for anything more complicated than a couple of particles becomes problematic even with modern computers. We need to dramatically simplify the problem. Fortunately, we can use some approximations to reduce the complexity of solving for the ground-state energy. Our approximations remove as many irrelevant degrees of freedom as possible.

One common approximation is to separate the nuclear and electronic degrees of freedom. Since the nuclei are considerably more massive than the electrons, we assume that the electrons will respond "instantaneously" to the nuclear coordinates. This approximation is called the *Born–Oppenheimer* or adiabatic approximation. It allows one to treat the nuclear coordinates as classical parameters. For most molecular systems, this assumption is highly accurate.

Another common approximation is to construct a specific form for the many-body wave function. If one can obtain an accurate estimate for the wave function, then from the variational principle an accurate estimate for the energy will emerge.

We can consider some simple examples to illustrate this approach. Suppose we consider a solution for *noninteracting electrons*, i.e. in Eq. (3.19), by ignoring the last term in the Hamiltonian. In this limit, one may write the many-body wave function as a sum of independent Hamiltonians. Using the adiabatic approximation, the *electronic* part of the Hamiltonian becomes

$$\mathscr{H}_{\text{el}}\left(\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots\right)=\sum_{i=1}^{M}\frac{-\hbar^2\nabla_i^2}{2m}-\sum_{n=1}^{N}\sum_{i=1}^{M}\frac{Z_n\text{e}^2}{|\vec{R}_n-\vec{r}_i|} \tag{3.22}$$

Let us define a nuclear potential, V_N, which the ith electron sees as

$$V_\text{N}(\vec{r}_i)=-\sum_{n=1}^{N}\frac{Z_n\text{e}^2}{|\vec{R}_n-\vec{r}_i|} \tag{3.23}$$

One can simplify the Schrödinger equation by writing

$$\mathscr{H}_{\text{el}}\left(\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots\right)\psi\left(\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots\right)=\sum_{i=1}^{M}H^i\psi\left(\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots\right) \tag{3.24}$$

where the electronic part of the Hamiltonian is now defined for the ith electron as

$$H^i=\frac{-\hbar^2\nabla_i^2}{2m}+V_\text{N}(\vec{r}_i) \tag{3.25}$$

For this simple Hamiltonian, let us write the many-body wave function as

$$\psi(\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots)=\phi_1(\vec{r}_1)\phi_2(\vec{r}_2)\phi_3(\vec{r}_3)\ldots \tag{3.26}$$

The $\phi_i(\vec{r})$ orbitals are determined from a "one-electron" Hamiltonian

$$\begin{aligned}H^i\phi_i(\vec{r})&=\left(\frac{-\hbar^2\nabla^2}{2m}+V_\text{N}(\vec{r})\right)\phi_i(\vec{r})\\&=E_i\phi_i(\vec{r})\end{aligned} \tag{3.27}$$

The index i for the orbital $\phi_i(\vec{r})$ is taken to include the spin of the electron plus any other relevant quantum numbers. The index i runs over the number of electrons, each electron being assigned a unique set of quantum numbers. This form of the Schrödinger equation is easily solved, more or less. The many-body wave function expressed in Eq. (3.26) is the *Hartree* wave function. The *Hartree Approximation*, named for Douglas Hartree, uses this form of the wave function to solve the Hamiltonian, *including* the electron–electron interactions [1]. By ignoring these interactions, the Hartree approximation assumes that the electrons are moving independently. The sum of the eigenvalues, E_i, gives the total energy of the system. If we wish to obtain a more realistic Hamiltonian, we must include electron–electron interactions, Eq. (3.24):

$$\mathscr{H}_{\text{el}}(\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots)\psi(\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots)=\sum_{i=1}^{M}\left(H^i+\frac{1}{2}\sum_{j=1,j\neq i}^{M}\frac{\text{e}^2}{|\vec{r}_i-\vec{r}_j|}\right)\psi(\vec{r}_1,\vec{r}_2,\vec{r}_3,\ldots) \tag{3.28}$$

We determine the individual orbitals, $\phi_i(\vec{r})$, by minimizing the total energy as per (Eq. (3.21)) with a constraint that the wave functions are normalized. Standard quantum mechanics books [2, 3] give the details of this procedure, which results in the following *Hartree* equation:

$$H^i\phi_i(\vec{r}) = \left(\frac{-\hbar^2\nabla^2}{2m} + V_N(\vec{r}) + \sum_{j=1,j\neq i}^{M} \int \frac{\mathrm{e}^2\,|\phi_j(\vec{r}\,')|^2}{|\vec{r}-\vec{r}\,'|}\,\mathrm{d}^3r\,' \right) \phi_i(\vec{r}) = E_i\phi_i(\vec{r}) \tag{3.29}$$

Using the orbitals, $\phi_i(\vec{r})$, from a solution of Eq. (3.29), we can construct a Hartree many-body wave function and find the total energy determined from (Eq. (3.21)).

The Hartree approximation is useful as an illustrative tool, but is a poor approximation to reality. It has a really significant flaw: The Hartree wave function fails to reflect the *antisymmetric* nature of the ectronic states, which is consistent with the Pauli Principle. Also, the Hartree equation is really difficult to solve. The Hamiltonian is orbitally dependent because the summation in Eq. (3.29) excludes the ith orbital. This means that if there are M electrons, then we need to solve the problem for M different Hamiltonians.

3.3 The Hartree–Fock Approximation

The symmetric nature of Hartree approximation results in a severely flawed many-body wave function. Owing to the work of John Slater, the antisymmetric nature of the wave function can be explicitly incorporated resulting in the Hartree–Fock approximation, named for the aforementioned Hartree and Vladimir Fock.

Here we need to consider the spin coordinate of each electron. The coordinates of an electron are specified by $\vec{r}_i s_i$ where s_i represents spin. Starting with one-electron orbitals, $\phi_i(\vec{r}s)$, we write a many-electron wave function as follows:

$$\psi\left(\vec{r}_1s_1, \vec{r}_2s_2, \vec{r}_3s_3, \dots\right) = \begin{vmatrix} \phi_1\left(\vec{r}_1s_1\right) & \phi_1\left(\vec{r}_2s_2\right) & \cdots & \cdots & \phi_1\left(\vec{r}_Ms_M\right) \\ \phi_2\left(\vec{r}_1s_1\right) & \phi_2\left(\vec{r}_2s_2\right) & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \phi_M\left(\vec{r}_1s_1\right) & \cdots & \cdots & \cdots & \phi_M\left(\vec{r}_Ms_M\right) \end{vmatrix} \tag{3.30}$$

This wave function as presented is called a *Slater determinant*. The determinant reflects the proper symmetry of the wave function and the Pauli principle. If two electrons occupy the same orbit, two rows of the determinant are identical, which sets the determinant to zero, and the many-body wave function vanishes. Likewise, the determinant will vanish if two electrons occupy the same point in generalized space (i.e. $\vec{r}_is_i = \vec{r}_js_j$) as two columns of the determinant are identical, again resulting in the determinant vanishing. If two electrons are exchanged, this corresponds to exchanging two columns and results in a sign change of the determinant. The Slater determinant is a very convenient representation as it automatically keeps track of the antisymmetric nature of the wave function.

If one uses a Slater determinant to evaluate the total electronic energy and maintains the orbital normalization, we can obtain the orbitals from the following *Hartree–Fock* equations:

$$H^i\phi_i(\vec{r}) = \left(\frac{-\hbar^2\nabla^2}{2m} + V_N(\vec{r}) + \sum_{j=1}^{M} \int \frac{\mathrm{e}^2\,|\phi_j(\vec{r}\,')|^2}{|\vec{r}-\vec{r}\,'|}\,\mathrm{d}^3r\,' \right)\phi_i(\vec{r})$$
$$-\sum_{j=1}^{M} \int \frac{\mathrm{e}^2}{|\vec{r}-\vec{r}\,'|}\,\phi_j^*(\vec{r}\,')\phi_i(\vec{r}\,')\,\mathrm{d}^3r\,'\delta_{s_i,s_j}\,\phi_j(\vec{r}) = E_i\phi_i(\vec{r}) \tag{3.31}$$

It is customary to simplify this expression by defining an electronic charge density, ρ:

$$\rho(\vec{r}) = \sum_{j=1}^{M} |\phi_j(\vec{r}\,)|^2 \tag{3.32}$$

and an orbitally dependent *exchange-charge density*, ρ_i^{HF} for the ith orbital:

$$\rho_i^{\mathrm{HF}}\left(\vec{r},\vec{r}\,'\right) = \sum_{j=1}^{M} \frac{\phi_j^*(\vec{r}\,')\,\phi_i(\vec{r}\,')\,\phi_i^*(\vec{r}\,)\,\phi_j(\vec{r}\,)}{\phi_i^*(\vec{r}\,)\,\phi_i(\vec{r}\,)}\,\delta_{s_i,s_j} \tag{3.33}$$

The exchange-charge density couples only states (i,j) with the same spin coordinates (s_i, s_j) and is dependent on $\vec{r},\vec{r}\,'$.

We can define the corresponding potential, the *Coulomb* or *Hartree* potential, V_{H}, by

$$V_{\mathrm{H}}(\vec{r}) = \int \rho(\vec{r})\,\frac{\mathrm{e}^2}{|\vec{r}-\vec{r}\,'|}\,\mathrm{d}^3r' \tag{3.34}$$

The *exchange* potential is defined by

$$V_x^i(\vec{r}) = -\int \rho_i^{\mathrm{HF}}(\vec{r},\vec{r}\,')\,\frac{\mathrm{e}^2}{|\vec{r}-\vec{r}\,'|}\,\mathrm{d}^3r' \tag{3.35}$$

This combination results in the following Hartree–Fock equation:

$$\left(\frac{-\hbar^2\nabla^2}{2m} + V_{\mathrm{N}}(\vec{r}) + V_{\mathrm{H}}(\vec{r}) + V_x^i(\vec{r}) \right)\phi_i(\vec{r}) = E_i\phi_i(\vec{r}) \tag{3.36}$$

Once the Hartree–Fock orbitals are obtained, the total Hartree–Fock electronic energy of the system, E_{HF}, is given by

$$E_{\mathrm{HF}} = \sum_i^M E_i - \frac{1}{2}\int \rho(\vec{r})V_{\mathrm{H}}(\vec{r})\,\mathrm{d}^3r - \frac{1}{2}\sum_i^M \int \phi_i^*(\vec{r}\,)\,\phi_i(\vec{r}\,)V_x^i(\vec{r})\,\mathrm{d}^3r \tag{3.37}$$

E_{HF} is *not* a sum of the Hartree–Fock orbital energies, E_i. The factor of $\frac{1}{2}$ in the Electron–electron terms arises because the electron–electron interactions have been double counted in the Coulomb and exchange potentials.

The Hartree–Fock Schrödinger equation is only slightly more complex than the Hartree equation. The Hartree and Hartree–Fock equations are difficult to solve because the exchange potential remains orbitally dependent, i.e. if we need to solve for M orbitals, we have M equations to solve. However, there is one notable difference in the Hartree–Fock summations compared to the Hartree summation. The Hartree–Fock sums include the $i = j$ terms in Eq. (3.31). This difference arises because the exchange term corresponding to $i = j$ cancels an equivalent term in the Coulomb summation. The $i = j$ term in both the Coulomb and exchange term is interpreted as a *self-screening* of the electron. In the Hartree formalism the electron is explicitly excluded from the screening potential and in the Hartree–Fock formalism the electron self-screening in the Coulomb potential is exactly cancelled by the self-screening term in the exchange potential. Without such an exact cancellation between the Coulomb and exchange terms, a "self-energy" contribution to the total energy will occur. Approximate forms of the exchange potential often do not have this property. The total energy then contains a self-energy contribution that one needs to remove to obtain a correct Hartree–Fock energy.

The Hartree–Fock wave functions are only *approximations* to the true ground-state many-body wave functions. Energy contributions absent in the Hartree–Fock approximation are referred to as *correlation* contributions. One definition for the "correlation energy," E_{corr}, is to write it as the difference between the correct total energy of the system and the Hartree–Fock energies: $E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}$. Feynman used a slightly less than "scientific" nomenclature for the correlation energy. He referred to it as the "stupidity energy" because it indicates how difficult it is to determine this quantity [4]. This is an accurate characterization. While more accurate computations are available with contemporary high-performance computers, calculations for the stupidity energy remain challenging.

To mitigate our "stupidity," we consider free electrons moving within a box where the electrons do not interact. We can solve this system "exactly" in the next chapter.

References

1 Hartree, D.R. (1957). *The Calculation of Atomic Structures.* Wiley.

2 Schiff, L. (1968). *Quantum Mechanics*, 3e. McGraw-Hill.

3 Shankar, R. (1994). *Principles of Quantum Mechanics.* Springer Science.

4 Feynman, R.P. (1972). *Statistical Mechanics.* Addison-Wesley.

4

The Free Electron Gas

> *All that glisters may not be gold, but at least it contains free electrons. [But consider the Golden Scarab Beetle which has a metallic luster without metal.]*
>
> — John Desmond Bernal

4.1 Free Electrons

How should we think about the exchange potential? Can it be simplified to make "practical computations"? Our goal in this chapter is to consider a picture where we have hope of extracting the essential features of this potential without the corresponding complications.

For the moment, we consider one electron contained in a box. The Schrödinger equation for this system is similar to Eq. (2.11) with the potential set to zero.

$$\frac{-\hbar^2\nabla^2}{2m}\,\phi_{\vec{k}}(\vec{r}) = E_{\vec{k}}\,\phi_{\vec{k}}(\vec{r}) \tag{4.1}$$

Ignoring spin for the moment, the solution of Eq. (4.1) is

$$\phi_{\vec{k}}(\vec{r}) = \frac{1}{\sqrt{\Omega}}\,\exp(i\vec{k}\cdot\vec{r}) \tag{4.2}$$

The energy is given by $E_{\vec{k}} = \hbar^2k^2/2m$, where $\vec{k}$ is called a *wave vector*. Since the quantum operator for momentum is $\vec{p} = -i\hbar\nabla$, the wave vector can be related to the electron's momentum by $\vec{p} = \hbar\vec{k}$. We can consider a Cartesian system and write $\vec{k} = \{k_x, k_y, k_z\}$: $\vec{k}$ is a set of three quantum numbers. $\vec{k}$ is called a good quantum number as one can label the wave functions by it. If we were to consider spin, we would have four quantum numbers, which is the same number as that for the hydrogen atom. We consider the wave function to exist over some volume, Ω, for normalizing the wave function. Suppose we have a hydrogen atom and change the proton's Coulomb field, which binds the electron, to a box that contains the electron. The quantum numbers for the hydrogen atom, n, l, m_l, m_s, will be"transmuted" to k_x, k_y, k_z, m_s. We normalize the wave function for our electron in a box as follows:

$$\int_\Omega \phi^*{}_{\vec{k}}(\vec{r})\phi_{\vec{k}}(\vec{r})\,\mathrm{d}^3r = \frac{1}{\Omega}\int_\Omega \exp(-i\vec{k}\cdot\vec{r})\exp(i\vec{k}\cdot\vec{r})\,\mathrm{d}^3r = \frac{1}{\Omega}\int_\Omega \mathrm{d}^3r = 1 \tag{4.3}$$

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

The integral is over the volume of the box and the normalization to unity means that the electron exists somewhere in the box. The probability density or charge density is given by $\rho = \phi^*_{\vec{k}}\, \phi_{\vec{k}} = 1/\Omega$, which means the electron is equally likely to be anywhere within the containing box. Formally, this must be true if space is isotropic – one point in space is as good as any.

This model for the electron is all right as is. Still, there is an annoying complication: A box has a surface or a boundary and we would like to avoid dealing with this issue. Without loss of generality, we consider a cubic box of size $L \times L \times L$ and take the origin at (0,0,0). We impose an infinitely large potential that will confine the electron to reside within the box. The wave function must vanish outside this box. Otherwise, the electron would experience the infinite potential outside the box and have infinite energy.

We rewrite the wave function to avoid such an event. Within the box we have free electron solution:

$$\phi_{\vec{k}}(x, y, z) = A \exp(i\vec{k} \cdot \vec{r}) + B \exp(-i\vec{k} \cdot \vec{r}) \tag{4.4}$$

and outside the box this wave function will vanish. Along the x-axis, $\phi_{\vec{k}}(0, y, z) = 0$ and $\phi_{\vec{k}}(L, y, z) = 0$; otherwise, our wave function would be discontinuous at the boundary of the box. We can separate out the x-axis behavior:

$$\phi_{\vec{k}}(0, y, z) = A + B = 0 \quad ; \quad \phi_{\vec{k}}(L, y, z) = A \exp(ik_x L) + B \exp(-ik_x L) = 0 \tag{4.5}$$

The first condition demands $A = -B$, so we can express the second condition as

$$\sin(k_x L) = 0 \qquad k_x L = \pi, 2\pi, 3\pi, \dots, n_x\pi, \dots \tag{4.6}$$

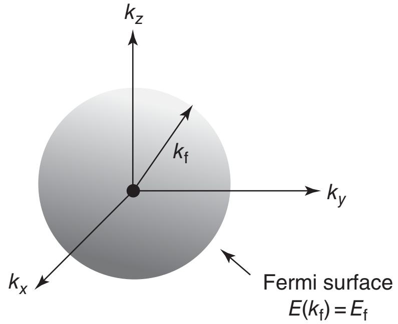
where n_x is a positive integer. Of course, we have the same condition for k_y, k_z and this means that we can write the energy of the electron as

$$E_{\vec{k}} = \frac{\hbar^2(k_x^2 + k_y^2 + k_z^2)}{2m} = \frac{\hbar^2\pi^2}{2mL^2}(n_x^2 + n_y^2 + n_z^2) \tag{4.7}$$

The boundary conditions require that $k = |\vec{k}|$ assume discrete values. This is expected for a quantum solution. However, unlike the case of a hydrogen atom, where the discrete values of the energy can be separated by electronvolts, the separation in values here will scale as $1/L^2$. We take L to a macroscopic size as we do not want the surface to play a significant role, i.e. the surface to volume ratio of the box will scale as $1/L$ and we are considering the limit of large L, in practice $L \to \infty$. To get a feel for the energy scales involved, we take $L = 1$ cm, then $E \sim \hbar^2\pi^2/2mL^2$ is about 10^{-15} eV, which is a very, very small number! An electron in such a huge box (relatively speaking) will not experience discrete energy levels, at least not on a scale that can be measured.

The problem we face is to consider many electrons in our box. We make the striking stipulation that the electrons do not interact. This system is called, logically, a *free electron gas*. This gas is unlike an ideal gas of atoms or molecules, e.g. air molecules, as the electrons are quantum particles carrying spins, *i.e.* Fermions, and they must obey the Pauli principle.

Figure 4.1 Fermi surface for a free electron gas.



How can we quantify the number and occupation of these states? Suppose we put 10^{23} electrons in this box. We decidedly do not want to inventory them. Instead, we can define a *density of states* that will tell us how many states exist per energy interval and we then occupy the energy levels as per Pauli.

We define a density of states, $D(E)$, as

$$D(E) = \frac{N(E + \Delta E) - N(E)}{\Delta E} \tag{4.8}$$

where $N(E)$ gives all the states below an energy E. Here, we have so many states that we can treat E as a continuous variable and take the limit of $\Delta E \to 0$. In this case,

$$D(E) = \frac{\mathrm{d}N}{\mathrm{d}E} \tag{4.9}$$

Our next job is to quantify N for a free electron gas. Since $E_{\vec{k}} = E(|\vec{k}|)$, we ask how many states are encompassed for a given k. We know that in $\{k_x, k_y, k_z\}$, or $\vec{k}$- space, a given value of k corresponds to a sphere as indicated in Figure 4.1. We fill in the electronic states with the lowest energy states first. The highest filled levels form a "Fermi surface," which in our case is the surface of a sphere. For free electrons, the relationship between E and k is trivial, $E_{\vec{k}} = \hbar^2 k/2m$. For electrons in real solids, this simple relationship may not hold as we will discuss later on.

How many states exist per unit volume? This is not a difficult question as each axis has allowed points separated by π/L. There is one allowed electron in a volume of $(\pi/L)^3$ as indicated in Figure 4.2. The total number of states, N, in the spherical volume is given by the volume of a sphere of radius k, divided by the volume per state:

$$N(k) = 2\,\frac{\text{volume of sphere}}{\text{volume per state}} = 2\,\frac{4\pi k^3/3}{8(\pi/L)^3} \tag{4.10}$$

where we included a factor of 2 from spin; the factor of 8 comes the fact that n_x, n_y, n_z are positive integers so we only consider 1/8 of the sphere. We can find $N(E)$ by replacing k by $k = \sqrt{2mE/\hbar^2}$:

$$N(E) = 2\,\frac{4\pi/3}{8(\pi/L)^3}\,(2mE/\hbar^2)^{3/2} \tag{4.11}$$

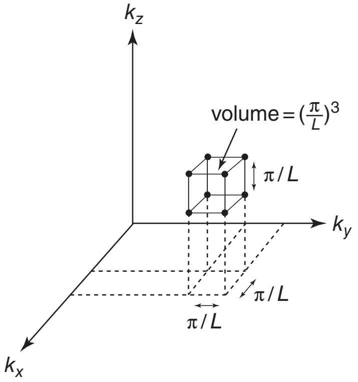


Figure 4.2 Example of volume per allowed value of (k_x, k_y, k_z). There are eight allowed points in the figure, but each point is shared by neighboring volumes. The volume per point is $(\pi/L)^3$ as shown in the figure.

The derivative of this expression gives us the number of states per unit energy, i.e. the density of states:

$$D(E) = \frac{L^3}{2\pi^2}\,(2m/\hbar^2)^{3/2}\,\sqrt{E} \tag{4.12}$$

We determine the energy of the highest occupied state, which is called the Fermi energy, E_f:

$$\mathcal{N} = \frac{L^3}{2\pi^2}\,(2m/\hbar^2)^{3/2}\int_0^{E_\text{f}}\sqrt{E}\,\text{d}E = \frac{L^3}{3\pi^2}\,(2m/\hbar^2)^{3/2}E_\text{f}^{3/2} \tag{4.13}$$

where $\mathcal{N}$ is the total number of electrons in the system. We can rewrite this defining the electron density as $n = \mathcal{N}/L^3$. We obtain the following:

$$E_\text{f} = \frac{\hbar^2}{2m}\,(3\pi^2 n)^{2/3} = \frac{\hbar^2 k_\text{f}^2}{2m} \tag{4.14}$$

where the Fermi wave vector, k_f, is defined by $k_\text{f} = (3\pi^2 n)^{1/3}$. Another useful relationship is

$$D(E_\text{f}) = \frac{3\mathcal{N}}{2E_\text{f}} \tag{4.15}$$

Within a numerical factor of order unity, the number of states at the Fermi energy is simply the total number of electrons in the box divided by the Fermi energy.

The density of states is used to find the total energy of the electrons, E_T, in our box:

$$E_\text{T} = \int_0^{E_\text{f}} E\,D(E)\,\text{d}E = \frac{L^3}{2\pi^2}\,(2m/\hbar^2)^{3/2}\int_0^{E_\text{f}} E^{3/2}\,\text{d}E = \frac{3}{5}\mathcal{N}\,E_\text{f} \tag{4.16}$$

The average energy per electron is 3/5 times the Fermi energy.

Given the "free electron" nature of the system, does this description have physical meaning, or is this an "academic exercise"? For many simple metals, i.e. metals with s and p valence states such as Na or Al, the theory works well, despite the extreme simplicity of the model. This is not a purely academic issue. We can use this model to compute several physical properties.

For example, one might argue that the increase in kinetic energy of the free electron gas under pressure dominates the bulk moduli of simple metal solids.

Table 4.1 Bulk moduli, *B*, for simple metals within the free electron model where *n* is the electron density, and E_f is the Fermi energy.

Metal	n (m^{-3} × 10^{28})	E_f (eV)	E_f (J×10^{-19})	B (theory) (N/m^2 × 10^{11})	B (experiment) (N/m^2 × 10^{11})
Li	4.7	4.72	7.55	0.236	0.116
Na	2.65	3.23	5.17	0.091	0.068
K	1.4	2.12	3.40	0.032	0.032
Rb	1.15	1.85	2.96	0.023	0.031
Mg	8.60	7.13	11.4	0.654	0.354
Ca	4.60	4.68	7.51	0.230	0.152
Sr	3.56	3.95	6.33	0.150	0.116
Ba	3.20	3.65	5.85	0.125	0.103

Source: Experiment as compiled by Kittel [1].

This would be true for an ideal gas, but we are not dealing with an ideal gas. Does it work for a free electron gas? We can compute the bulk modulus, B, from $B = -V(\partial p/\partial V)$, where p is the pressure and V is the volume. The pressure is defined by $p = -\partial E_T/\partial V$. Putting this together with the Eqs. (4.14) and (4.16) yields $B = 10E_T/9V = 2nE_f/3$ where n is the electron density. In Table 4.1, we compare predicted and measured values of the bulk modules for some alkali and alkaline earth metals [1].

The most notable errors occur for the light metals Li and Mg. The free electron gas model overestimates the bulk modulus by a factor of two. However, the agreement is quite good for metals such as K and Sr. Overall, the trend of the bulk modulus as a function of electron density replicates experiment and supports the notion that under pressure the kinetic energy of the electrons is responsible for the compressibilities of metals.

We illustrate another case where the free electron model does a reasonable job in describing the physical properties of metals. Consider the contribution of electrons to the heat capacity of a simple metal. Heat capacity relates the change in the internal energy of material to a change in temperature. For a metal, the internal energy change is related to the electronic states. Classically, we expect each degree of freedom to possess an energy of $k_B T/2$, where k_B is the Boltzmann constant. For a metal containing $\mathcal{N}$ electrons, we would expect the heat capacity to be $3\mathcal{N}k_B/2$. Classical physics gives this type of behavior, and it is all right for many solids at high temperatures where the heat capacity goes to a constant value. However, classical physics is clearly wrong for a free electron gas, especially at low temperatures where the measured heat capacity of a simple metal vanishes.

As an aside, the change in entropy of a solid with temperature is related to the heat capacity:

$$\Delta S = \int_0^T \frac{C}{T}\, dT \tag{4.17}$$

If C does not vanish as $T \to 0$, the change in entropy is not defined. Suppose $C \approx a_0 + a_1 T$.

$$\Delta S \approx a_0 \ln(T) + a_1 T - a_0 \ln(0) \tag{4.18}$$

with $\ln(0) \to -\infty$. The divergence of the entropy is unacceptable and $a_0 \neq 0$ is bad.

Quantum mechanics gives a different result and correctly accounts for heat capacities at small temperatures. Einstein was the first to recognize this issue, but not for a free electron gas. For metals, the argument goes like this.

If an electron changes its state because of thermal excitation, we expect it to change its $\vec{k}$-state. For an electron near the Fermi energy, this is not an issue. There are a number of empty states close in energy. However, consider an electron whose kinetic energy is much below the Fermi level, say its kinetic energy is zero. The Pauli principle prohibits the electron from occupying an already filled state. The electron needs to be excited to an empty state above the Fermi level. The energy required to move an electron in a simple metal such as potassium from a zero energy state to above the Fermi energy is roughly 2 eV (see Table 4.1). This is a lot of energy. It corresponds to a temperature of ~23 000 K – hotter than the surface of the sun and far above the melting point of K. That will not be the case when the temperature approaches 0 K.

In general, an electron will only be excited if it is in $k_B T$ of the Fermi energy. The number of electrons that can be excited at a temperature T can be estimated as follows:

$$\Delta\mathcal{N} \approx D(E)\, k_B T = \frac{3\mathcal{N}}{2E_f} k_B T \tag{4.19}$$

This gives an electronic heat capacity, C_{elec}, of the derivative of $3k_B T \Delta\mathcal{N}/2$

$$C_{\mathrm{elec}} = \frac{\mathrm{d}E}{\mathrm{d}T} \approx \frac{\mathrm{d}}{\mathrm{d}T} \frac{3\mathcal{N}(k_B T)^2}{4E_f} = 3\frac{\mathcal{N}(k_B)^2 T}{2E_f} \tag{4.20}$$

A more accurate treatment of the heat capacity yields

$$C_{\mathrm{elec}} = \frac{\mathrm{d}E}{\mathrm{d}T} = \pi^2 \frac{\mathcal{N}(k_B)^2 T}{2E_f} = \pi^2 \frac{\mathcal{N} k_B T}{2T_f} \tag{4.21}$$

where $E_f = k_B T_f$. Pauli principle limitations on low-lying excitations as the temperature approaches 0 K accounts for the difference between the classical and quantum descriptions of the heat capacity.

In Table 4.2 we compile the heat capacity constant $\gamma = \pi^2 \frac{\mathcal{N}_\dashv k_B}{2T_f}$ for some simple metals where $\mathcal{N}_\dashv$ is Avogadro's number. We note that the ratio: $\gamma_{\mathrm{meas}}/\gamma_{\mathrm{free}}$ exceeds unity. This ratio is often named the "thermal effective mass" to the electron mass, i.e. $m_{\mathrm{th}}/m = \gamma_{\mathrm{meas}}/\gamma_{\mathrm{free}}$. The mass ratio deviates from unity owing to several effects such as the interaction of electrons with true electronic potentials of the metal and the interaction of the electrons among themselves and with lattice vibrations.

Table 4.2 Heat capacity constant, γ, for selected simple metals compared to experiment. Units are in mJ mol^{-1} K^{-2}.

Metal	Free electron model	Experiment
Li	0.75	1.63
Na	1.09	1.38
K	1.67	2.08
Rb	1.91	2.41
Mg	0.99	1.3
Ca	1.51	2.9
Sr	1.79	3.6
Ba	1.94	2.7

Source: Experiment is as compiled by Kittel [1].

4.2 Hartree–Fock Exchange in a Free Electron Gas

The simplicity of the free electron gas model allows one to predict physical properties such as the heat capacity and compressibility of metals. So perhaps it can be used to yield insights on how electrons interact in "real materials." We build on this model to consider the "many-electron" problem. For example, it is possible to evaluate the Hartree–Fock exchange energy directly and analytically for a free electron gas.

Suppose we construct a Slater determinant using free electron orbitals. We label each orbital by $\vec{k}$ and include a spin index. The Coulomb potential for an infinite free electron gas diverges, but this divergence can be removed by imposing a compensating uniform positive charge. The combination of a free electron gas supported by a uniform positive background is the simplest model for a metal and is a called a "jellium model."

The resulting Hartree–Fock eigenvalues can be found, again using standard textbooks [2]:

$$E_k = \frac{\hbar^2 k^2}{2m} - \frac{1}{L^3} \sum_{k'<k_\mathrm{f}} \frac{4\pi e^2}{|\vec{k} - \vec{k}'|^2} \tag{4.22}$$

where the summation is over occupied $\vec{k}$-states. In principle, this means summing over (n_x, n_y, n_z). However, it is possible to evaluate the summation by transforming the summation into an integration. This transformation is easily performed if we consider a continuum of states:

$$\frac{1}{L^3} \sum_{k'<k_\mathrm{f}} \frac{4\pi e^2}{|\vec{k} - \vec{k}'|^2} = \frac{1}{(2\pi)^3} \int_{k'<k_\mathrm{f}} \frac{4\pi e^2}{|\vec{k} - \vec{k}'|^2} \, d^3k' \tag{4.23}$$

This integral can be solved analytically, although it looks difficult. The resulting eigenvalues are given by

$$E_k = \frac{\hbar^2 k^2}{2m} - \frac{e^2 k_\mathrm{f}}{\pi} \left(1 + \frac{1 - (k/k_\mathrm{f})^2}{2(k/k_\mathrm{f})} \ln \left| \frac{k + k_\mathrm{f}}{k - k_\mathrm{f}} \right| \right) \tag{4.24}$$

Using the above expression and Eq. (3.37), the total electron energy, $E_{\mathrm{HF}}^{\mathrm{FEG}}$, for a free electron gas within the Hartree–Fock approximation is given by

$$E_{\mathrm{HF}}^{\mathrm{FEG}} = 2\sum_{k<k_{\mathrm{f}}} \frac{\hbar^2 k^2}{2m} - \frac{\mathrm{e}^2 k_{\mathrm{f}}}{\pi} \sum_{k<k_{\mathrm{f}}} \left(1 + \frac{1-(k/k_{\mathrm{f}})^2}{2(k/k_{\mathrm{f}})} \ln\left|\frac{k+k_{\mathrm{f}}}{k-k_{\mathrm{f}}}\right|\right) \tag{4.25}$$

The factor of 2 in the first term comes from spin. In the exchange term, there is no extra factor of 2 because one must substract off a "double counting term" (see Eq. (3.37)). The summations can be executed as per Eq. (4.23) to yield

$$E_{\mathrm{HF}}^{\mathrm{FEG}}/\mathcal{N} = \frac{3}{5}E_{\mathrm{f}} - \frac{3\mathrm{e}^2}{4\pi}k_{\mathrm{f}} \tag{4.26}$$

The first term corresponds to the average energy per electron in a free electron gas as we found earlier. The second term is new and corresponds to the exchange energy per electron. The exchange energy is attractive. There is a physical argument for its sign. One can argue that the spin of electrons can act to reduce their electronic energy. Consider two electrons with the same spin. Quantum theory insists that such electrons not be at the same point in space. (Electrons cannot have the same quantum number and be at the same point in space.) If that were to happen, the Slater determinant for the electronic many-body wave function would vanish. Even without a repulsive Coulomb interaction electrons with the same spin will avoid each other. This additional "repulsion" reduces the Coulomb interaction and makes the exchange interaction attractive.

The exchange energy term has a factor of k_{f}, which scales as the cube root of the free electron gas. Although our use of a free electron gas is slightly simplified, the resulting form for the exchange potential provides a clue as to what form the exchange energy might take in an interacting electron gas or nonuniform electron gas.

References

1 Kittel, C. (2005). *Introduction to Solid State Physics*, 8e. Wiley, contains tables of date for the free electron gas model.

2 Kittel, C. (1987). *Quantum Theory of Solids*. Wiley.

5

Density Functional Theory

> *In general the many-electron wave function,*$\Psi(r_1, r_2, \dots, r_n)$*, for a system of n electrons is not a legitimate scientific concept, when n exceeds ~1,000.*
>
> — Walter Kohn

5.1 Thomas–Fermi Theory

Given the complexity of the many-electron problem, there have been a number of attempts to simplify the problem and avoid a direct solution of the Schrödinger equation. Thomas and Fermi independently proposed a procedure based on the electronic charge density [1, 2]. They argued that knowing the electronic density obviates the need to solve for wave functions. Their approach has great merit. In quantum theory, we cannot directly measure wave functions as they are not observable in quantum mechanics. Yes, this is an odd situation. One of the chief components of quantum mechanics (the wave function) allows us to predict where an electron is, but only if we consider its modulus square wherein any phase dependence is removed. It is possible to detect relative differences of the wave function phase, e.g. in quantum effects such as the Josephson effect or in science fiction shows where universes differ by phase factors. This is an area where no one need go. Rather, one can consider the spatial distribution of electrons – the electron density – to be a fundamental quantity by itself and focus on computing it.

Thomas and Fermi proposed that the total electronic energy of a system such as an atom be written as

$$E_{\mathrm{T}} = \frac{3}{10}\frac{\hbar^2(3\pi^2)^{2/3}}{m}\int [n(\vec{r})]^{5/3}\,\mathrm{d}^3r + \int V_{\mathrm{ext}}\,n(\vec{r})\,\mathrm{d}^3r + \int n(\vec{r})V_{\mathrm{H}}(\vec{r})\,\mathrm{d}^3r \tag{5.1}$$

where $n(\vec{r})$ is the electron density, V_{ext} is a fixed external potential (such as the nuclear potential in an atom), and V_{H} is the Hartree potential as defined earlier. The first term in Eq. (5.1) is the kinetic energy; the second term corresponds to the electronic interaction with an external potential, which does not depend on the electronic density; and the third term is the electrostatic interaction, which

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

does. We write the electrostatic potential with the Hartree potential:

$$V_{\mathrm{H}}(\vec{r}) = \int \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|}\, \mathrm{d}^3 r' \tag{5.2}$$

The total electronic energy is minimized to determine a solution with an obvious constraint: The number of electrons in the system is conserved.

$$\mathcal{N} = \int n(\vec{r}) \mathrm{d}^3 r \tag{5.3}$$

where $\mathcal{N}$ is the total number of electrons in the system. We can solve for the density by finding the energy minimum, i.e. we find the best density by taking a "functional derivative" of the energy with respect to the density. What is a functional derivative? Here is an example. We write the electrostatic energy as follows:

$$E_{\mathrm{H}}[n] = \frac{1}{2} \iint \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|} \mathrm{d}^3 r\, \mathrm{d}^3 r' \tag{5.4}$$

$E_{\mathrm{H}}[n]$ is a functional of n. We put in a value of $\vec{r}$ and get n, which in turn gives the value of E_{H}. Suppose we vary the value of the function, n, and ask what the functional derivative of E_{H} is. The functional derivative can be extracted from

$$\int \frac{\delta E_{\mathrm{H}}}{\delta n} \phi(\vec{r})\, \mathrm{d}^3 r = \lim_{\varepsilon \to 0} \frac{E_{\mathrm{H}}[n + \varepsilon\phi] - E_{\mathrm{H}}[n]}{\varepsilon} \tag{5.5}$$

where $\delta E_{\mathrm{H}}/\delta n$ is the functional derivative that we want and ϕ is an arbitrary function. Using our definition of E_{H} we have

$$\int \frac{\delta E_{\mathrm{H}}}{\delta n} \phi(\vec{r})\, \mathrm{d}^3 r = \left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \frac{1}{2} \iint \frac{[n(\vec{r}) + \varepsilon\phi(\vec{r})][n(\vec{r}') + \varepsilon\phi(\vec{r}')]}{|\vec{r} - \vec{r}'|} \mathrm{d}^3 r\, \mathrm{d}^3 r' \right]_{\varepsilon \to 0} \tag{5.6}$$

which can be written as

$$\int \frac{\delta E_{\mathrm{H}}}{\delta n} \phi(\vec{r})\, \mathrm{d}^3 r = \frac{1}{2} \iint \frac{n(\vec{r}')\phi(\vec{r})}{|\vec{r} - \vec{r}'|} \mathrm{d}^3 r\, \mathrm{d}^3 r' + \frac{1}{2} \iint \frac{n(\vec{r})\phi(\vec{r}')}{|\vec{r} - \vec{r}'|} \mathrm{d}^3 r\, \mathrm{d}^3 r' \tag{5.7}$$

The two integrals on the right-hand side of the equation are equal as $\vec{r}$ and $\vec{r}'$ can be interchanged. This yields

$$\int \frac{\delta E_{\mathrm{H}}}{\delta n} \phi(\vec{r})\, \mathrm{d}^3 r = \int \left(\int \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} \mathrm{d}^3 r' \right) \phi(\vec{r})\, \mathrm{d}^3 r \tag{5.8}$$

We are done – the functional derivate is given by

$$\frac{\delta E_{\mathrm{H}}}{\delta n} = \int \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} \mathrm{d}^3 r' = V_{\mathrm{H}}(\vec{r}) \tag{5.9}$$

Note how the functional derivative of the energy turned into a potential. Consider the following functional derivative:

$$\frac{\delta}{\delta n} \left[E_{\mathrm{T}} + \lambda(\mathcal{N} - \int n(\vec{r}) \mathrm{d}^3 r) \right] = 0 \tag{5.10}$$

where λ is a Lagrange multiplier recognizing that we need to conserve the total number of electrons. Before we do something so bold, let us consider a trivial example. We ignore the Hartree potential and the external potential and assume that we have a free electron gas. In this limiting case, we make use of

$$\frac{\delta}{\delta n}\frac{3}{10}\frac{\hbar^2(3\pi^2)^{2/3}}{m}\int[n(\vec{r})]^{5/3}\,\mathrm{d}^3r=\frac{1}{2}\frac{\hbar^2(3\pi^2)^{2/3}}{m}[n(\vec{r})]^{2/3} \tag{5.11}$$

which yields

$$\frac{\hbar^2(3\pi^2)^{2/3}}{2m}[n(\vec{r})]^{2/3}-\lambda=0 \tag{5.12}$$

We recognize this expression as

$$\lambda=\frac{\hbar^2(3\pi^2n)^{2/3}}{2m}=E_\mathrm{f} \tag{5.13}$$

the Fermi energy for a free electron gas, i.e. the Lagrange multiplier is simply the Fermi energy, and the electron density is a constant, i.e. $n=\mathcal{N}/\Omega$ where Ω is the volume of the system. We can put this back into our total energy expression:

$$E_\mathrm{T}=\frac{3}{5}E_\mathrm{f}\int n\,\mathrm{d}^3r=\frac{3}{5}\mathcal{N}E_\mathrm{f} \tag{5.14}$$

and confirm that the total electronic energy is that of a free electron gas. Of course, we cheated! The kinetic energy expression was specifically chosen to be consistent with a free electron gas, but at least we verified it.

We should reinstate the external potential and the electrostatic potential, and identify λ with E_f. We arrive at the following expression for the electron density:

$$\frac{\hbar^2(3\pi^2)^{2/3}}{2m}[n(\vec{r})]^{2/3}+V_\mathrm{ext}+\int\frac{n(\vec{r}')}{|\vec{r}-\vec{r}'|}\,\mathrm{d}^3r'=E_\mathrm{f} \tag{5.15}$$

where we wrote out the V_H term to illustrate the presence of the electron density. This expression is decidedly easier to solve as compared to the Hartree–Fock description, e.g. we have no eigenvalue problem to solve. Owing to the simplicity of Thomas–Fermi theory, it led to the first quantitative attempt at describing the electronic density of an atom and its corresponding energy.

As usual, there is good news and bad news. First, we give the good news. The overall electronic density profile for an atom is reasonably well represented by Thomas–Fermi theory when compared to a quantum mechanical solution. The rest is all bad! There is no shell structure for a Thomas–Fermi atom as there are no orbitals, i.e. no s-states or p-states or d-states. The situation gets worse because there is also no bonding between Thomas–Fermi atoms! According to the Thomas–Fermi theory, no molecules are allowed to form from atoms. This was demonstrated by Teller [3].

Thomas–Fermi theory fails in part because the expression for the kinetic energy is simply awful for real atoms. Also, there is no explicit treatment of the exclusion principle in Thomas–Fermi theory as there are no orbitals. As such, the exchange energy resulting from the exclusion principle is absent and does not provide the attractive energy we found in the Hartree–Fock solution of free electrons.

Given such deficiencies, one might have thought any interest in density functional methods would be of academic interest. However, the lure of finding a simpler, more direct solution of the electronic structure problem is seductive and irresistible. And, progress was made by workers such as Slater and Gaspar who successfully implemented density functional theories into a quantum formalism [4, 5]. They did not replace the kinetic energy by a functional of the charge density, which we know is bad. Rather, they introduced an exchange term based on the electron density. They were able to solve the electronic structure problem for atoms, molecules, and solids with many notable successes.

Their approach did lack a fundamental theoretical framework and some leaders in the field viewed their work with some skepticism to say the least. For example, Ashcroft and Mermin wrote the following in their highly popular textbook [6] on condensed matter: "This procedure [using a free electron expression for the exchange energy] gross and *ad hoc* though it is, is *actually* followed in many band structure computations." [Emphasis added on "actually."] In some sense, they are correct. The method is and was overly simple. However, good science often starts with simple approaches. Also, it is always better to take a complicated problem and make it simpler, not more complicated. One wonders if Ashcroft and Mermin would have written such a negative view of density functional theory had they known this approach would eventually be honored with a Nobel Prize in 1998, 20 years after their textbook was published.

5.2 The Kohn–Sham Equation

In a set of seminal papers in the mid-1960s, Pierre Hohenberg, Walter Kohn, and Lu Sham established the theoretical framework for justifying the replacement of the many-body wavefunction by one-electron orbitals with the electron density being a key quantity [7, 8]. Hohenberg and Kohn argued that the total electronic energy of a system depended in a unique way on the electron density, with Sham and Kohn providing a recipe for computing the requisite density. For this work, Walter Kohn received the Chemistry Nobel Prize.

We outline some of the essential features in the Hohenberg–Kohn–Sham theory. Recall from our earlier discussion that a key problem in solving the Hartree–Fock equation

$$\left(\frac{-\hbar^2 \nabla^2}{2m} + V_{\mathrm{N}}(\vec{r}) + V_{\mathrm{H}}(\vec{r}) + V_x^i(\vec{r}) \right) \phi_i(\vec{r}) = E_i \phi_i(\vec{r}) \tag{5.16}$$

is the orbital dependence on the exchange term. If we have five occupied orbitals, then we would have five coupled Hamiltonians. Suppose for a moment that the exchange potential did not depend on the orbital, i.e. suppose $V_x^i = V_x$. The removal of the orbital dependence results in an immediate simplification by reducing the problem from five Hamiltonians to one. Kohn and Sham accomplished this simplification by the introduction of the *local density approximation*. Within this approximation, one expresses the exchange energy of the system as

$$E_x[n(\vec{r})] = \int n(\vec{r})\, \mathscr{E}_x[n(\vec{r})]\, \mathrm{d}^3 r \tag{5.17}$$

where $\mathscr{E}_x[n]$ is the "exchange energy density" for an electronic density of n. Within the Kohn–Sham framework, the exchange potential is determined from the functional derivative of $E_x[n]$:

$$V_x[n] = \frac{\delta E_x[n]}{\delta n} \tag{5.18}$$

Of course, this is a formal discussion without a specification of $\mathscr{E}_x[n]$, the exchange energy density. How does one obtain this quantity? We considered a limiting case – the free electron gas – in anticipation of this question. For a free electron gas, we know the exchange energy per electron:

$$\mathscr{E}_x = -\frac{3e^2}{4\pi}k_\mathrm{f} = -\frac{3e^2}{4\pi}(3\pi^2 n)^{1/3} \tag{5.19}$$

Suppose we make a bold assumption. Namely, we assume that this is a *universal* functional, which is generally applicable to all electronic structure problems – whether the density is uniform or not. Once we make this assumption, we can write

$$E_x[\rho] = -\frac{3e^2}{4\pi}(3\pi^2)^{1/3}\int [n(\vec{r})]^{4/3}\,\mathrm{d}^3r \tag{5.20}$$

Taking the functional derivative, we obtain

$$V_x[n] = -\frac{e^2}{\pi}(3\pi^2 n(\vec{r}))^{1/3} \tag{5.21}$$

This expression differs by a factor of 2/3 from the one proposed by Slater some 15 years before the work of Kohn and Sham. (Slater did not minimize the energy with respect to the free electron density; he did an averaging.) For a number of years, some controversy existed as to whether the Kohn–Sham or Slater exchange was more accurate for realistic systems [6].

Slater suggested an "engineering" solution to this issue. He introduced a parameter that could be varied between the original Slater and Kohn–Sham values [9]. The parameter, α, was often placed in front of the Slater exchange: $V_{x\alpha} = \alpha V_x^{\mathrm{Slater}}$. The α parameter was then chosen to replicate some known feature of an exact Hartree–Fock calculation such as the total energy of an atom or ion. Acceptable values of α were viewed to range from $\alpha = 2/3$ to $\alpha = 1$. Slater's so called "X_α" method was very successful in describing molecular systems.

Notable drawbacks of the X_α method center on its *ad hoc* nature through the α parameter and the omission of an explicit treatment of correlation energies. In contemporary theories, α is taken to be 2/3, and correlation energies are explicitly included in the energy functionals. Numerical studies have been performed on electron gases resulting in local density expressions of the form: $V_{\mathrm{xc}}[n(\vec{r})] = V_x[n(\vec{r})] + V_\mathrm{c}[n(\vec{r})]$ where V_c represents contributions to the total energy beyond the Hartree–Fock limit.

Local density functionals can also be improved by considering the role of spin. Suppose we consider different components of the charge density: one for spin up and one for spin down. We can write the total density as follows: $\rho = \rho_\uparrow + \rho_\downarrow$. This approximation is called the *local spin density approximation*. Functionals using this approximation treat "up" and "down" spins differently.

Given a functional for exchange and correlation, we can solve the *Kohn–Sham equation* for the electronic structure of matter:

$$\left(\frac{-\hbar^2 \nabla^2}{2m} + V_{\mathrm{N}}(\vec{r}) + V_{\mathrm{H}}(\vec{r}) + V_{\mathrm{xc}}[n(\vec{r})] \right) \phi_i(\vec{r}) = E_i \phi_i(\vec{r}) \tag{5.22}$$

This equation is solved “self-consistently.”

Here is how the method works: Initially, an approximate charge is assumed to estimate the exchange–correlation potential and this charge is used to determine the Hartree potential from Eq. (3.34). These approximate potentials are inserted in the Kohn–Sham equation and the total charge density determined as in Eq. (3.32). The “output” charge density is used to construct new exchange-correlation and Hartree potentials. The process is repeated until the input and output potentials are identical to within some specified tolerance.

The generation of a self-consistent solution of the Kohn–Sham equation is not quite as simple as it sounds. We need a reasonable charge density to initiate the process. This is not difficult. We can solve the Kohn–Sham equation for an isolated atom by assuming that the charge density of the atom is spherically symmetric. In this case, the problem becomes one dimensional and the resulting Kohn–Sham problem can readily be solved. Programs using Slater’s theory for solving for self-consistent atoms date back to the early 1960s with the invention of digital computers.

Once a solution of the Kohn–Sham equation is obtained, the total energy can be computed from

$$E_{\mathrm{KS}} = \sum_i^M E_i - 1/2 \int \rho(\vec{r}) V_{\mathrm{H}}(\vec{r})\, \mathrm{d}^3 r + \int \rho(\vec{r}) (\mathscr{E}_{\mathrm{xc}}[\rho(\vec{r})] - V_{\mathrm{xc}}[\rho(\vec{r})])\, \mathrm{d}^3 r \tag{5.23}$$

The electronic energy as determined from E_{KS} must be added to the ion–ion interactions to obtain the structural energies. This is a straightforward calculation for confined systems. For extended systems such as crystals, the calculations can be complex as individually the positive and negative charge distributions have Coulombic energies that individually diverge. Such summations can be done using Madelung techniques [6].

Owing to its ease of implementation and overall accuracy, local density approximation is the current method of choice for describing the electronic structure of condensed matter and is often implemented in the chemistry community. Thousands of papers have been published with this technique and its use appears to be increasing, despite some alternative approaches.

Some history: When first proposed, density functional theory was not widely accepted in the chemistry or physics community. We noted this earlier by quoting from the Aschroft and Mermin text. The theory is not “rigorous” in the sense that it is not clear how the estimates for the ground-state energies can be improved. For wave function based methods, one can include more Slater determinants as in a configuration interaction approach. Given more degrees of freedom, the trial wave functions improve the estimate for the electronic energy via the variational theorem. In density functional theory, there is no analogous procedure. The Kohn–Sham equations are variational, but need not approach the true ground-state energy as the functionals implemented to date are never exact. Obtaining an incorrect value of the absolute electronic energy need not be a problem, provided we are interested in *relative* energies, which we often are. For example, we might take the energy of two different structures to see which one is more stable. In such situations, the absolute energy is not of interest. Moreover, inherent density functional errors can cancel in taking the difference.

Error cancellation is thought to account for the utility of density functional theory. As an example, if the absolute energy of an atom is 1000 eV and the error in determining the energy is 100 eV (a 10% error), then it might seem unreasonable to compute a binding energy within 10 eV. However, if everything is consistently off by 10%, getting a binding energy of 10 eV is doable. An outstanding fundamental issue of using density functional theory is obtaining a reliable estimate of the cancellation of errors before doing the computation. To put it more succinctly, how do we know if density functional theory will work?

In some sense, density functional theory is an *a posteri* theory although it is often called an *ab initio* theory, i.e. a theory not dependent on experimental input. Given the transference of the exchange-correlation energies from an electron gas, it is not surprising that errors would arise in its implementation for highly nonuniform electron gas systems as found in realistic systems. However, the degree of error cancellations is rarely known *a priori*. The reliability of density functional theory has only been established by numerous calculations (literally thousands) for a wide variety of condensed matter and molecular systems. In short, density functional theory has proved itself in "combat" to be a practical approach for the quantum theory of materials.

References

1 Thomas, L.H. (1927). The calculation of atomic fields. *Proceedings of the Cambridge Philosophical Society* 23: 542.
2 Fermi, E. (1927). Un Metodo Statistico per la Determinazione di alcune Prioprietá dell'Atomo. *Rendiconti della Accademia Nazionale dei Lincei* 6: 602.
3 Teller, E. (1962). On the stability of molecules in the Thomas-Fermi theory. *Review of Modern Physics* 34: 627.
4 Slater, J.C. (1951). A simplification of the Hartree-Fock method. *Physical Review* 81: 385.
5 Gáspár, R. (1954). About an approximation of the Hartree-Fock potential by a universal potential function. *Acta Physica Academiae Scientiarum Hungaricae* 3: 263.
6 Aschroft, N.W. and Mermin, N.D. (1976). *Solid State Physics*, 1e. Brooks Cole.
7 Hohenberg, P. and Koh, P. (1964). Inhomogeneous electron gas. *Physical Review* 136: B864.
8 Kohn, W. and Sham, L.J. (1965). Self-consistent equations including exchange and correlation effects. *Physical Review* 140: A1133.
9 Slater, J.C. (1968). *Quantum Theory of Matter*, 2e. McGraw-Hill.

6

Pseudopotential Theory

> *The principles of an approximation method have been recently published by the author for a system consisting of closed shells and of valence electrons.*
>
> — Hans Hellmann

6.1 The Pseudopotential Approximation

Not all electronic states are created equal. Consider the carbon atom. The electronic configuration is $1s^22s^22p^2$. The 1s state is very tightly bound and highly localized around the nucleus. Within the Bohr model, the orbital size of an electron scales as $\sim 1/Z$ where Z is the number of protons within the nucleus. For a single electron orbiting the carbon nucleus, the Bohr model would give an orbital radius for a 1s state about 1/6 times that of the hydrogen atom or roughly 0.1 Å. In contrast, a carbon–carbon molecule has a bond length of roughly 1.5 Å, an order of magnitude larger than the 1s radius. Similarly, the binding energy of an electron in the 1s state in a carbon atom is more than an order of magnitude larger than that of a valence electron.

Moreover, the 1s state is not strongly altered when carbon atoms form bonds. Rather, a combination of the 2s and 2p states is known to form a chemical bond. For example, in the methane molecule, CH_4, the 2s and 2p states of the C atom hybridize to form a tetrahedral bonding configuration known as sp^3 bonding. Another example of the role of the valence electronic states occurs in simple metals such as sodium, which has an atomic configuration of $1s^22s^22p^63s^1$. The $1s^22s^22p^6$ electronic states are tightly bound and do not contribute to metallic bonds in elemental solid sodium. Only the $3s^1$ electron moves through the sodium crystal to carry current.

This dichotomy between tightly bound electronic states, called "core states," and the outermost, loosely bound electronic states, called "valence states," is recognized in the periodic table. Elements are grouped in columns based on the valence states, as shown in Figure 6.1.

Can one incorporate this physical idea of valence and core states into a workable approximation to help us solve the Kohn–Sham equation? Typically, the best numerical algorithms for solving physical problems are based on physical concepts.

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

hydrogen 1 H 1.0079																		helium 2 He 4.0026
lithium 3 Li 6.941	beryllium 4 Be 9.0122												boron 5 B 10.811	carbon 6 C 12.011	nitrogen 7 N 14.007	oxygen 8 O 15.999	fluorine 9 F 18.998	neon 10 Ne 20.180
sodium 11 Na 22.990	magnesium 12 Mg 24.305												aluminium 13 Al 26.982	silicon 14 Si 28.086	phosphorus 15 P 30.974	sulfur 16 S 32.065	chlorine 17 Cl 35.453	argon 18 Ar 39.948
potassium 19 K 39.098	calcium 20 Ca 40.078		scandium 21 Sc 44.956	titanium 22 Ti 47.867	vanadium 23 V 50.942	chromium 24 Cr 51.996	manganese 25 Mn 54.938	iron 26 Fe 55.845	cobalt 27 Co 58.933	nickel 28 Ni 58.693	copper 29 Cu 63.546	zinc 30 Zn 65.39	gallium 31 Ga 69.723	germanium 32 Ge 72.61	arsenic 33 As 74.922	selenium 34 Se 78.96	bromine 35 Br 79.904	krypton 36 Kr 83.80
rubidium 37 Rb 85.468	strontium 38 Sr 87.62		yttrium 39 Y 88.906	zirconium 40 Zr 91.224	niobium 41 Nb 92.906	molybdenum 42 Mo 95.94	technetium 43 Tc [98]	ruthenium 44 Ru 101.07	rhodium 45 Rh 102.91	palladium 46 Pd 106.42	silver 47 Ag 107.87	cadmium 48 Cd 112.41	indium 49 In 114.82	tin 50 Sn 118.71	antimony 51 Sb 121.76	tellurium 52 Te 127.60	iodine 53 I 126.90	xenon 54 Xe 131.29
caesium 55 Cs 132.91	barium 56 Ba 137.33	57-70 *	lutetium 71 Lu 174.97	hafnium 72 Hf 178.49	tantalum 73 Ta 180.95	tungsten 74 W 183.84	rhenium 75 Re 186.21	osmium 76 Os 190.23	iridium 77 Ir 192.22	platinum 78 Pt 195.08	gold 79 Au 196.97	mercury 80 Hg 200.59	thallium 81 Tl 204.38	lead 82 Pb 207.2	bismuth 83 Bi 208.98	polonium 84 Po [209]	astatine 85 At [210]	radon 86 Rn [222]
francium 87 Fr [223]	radium 88 Ra [226]	89-102 * *	lawrencium 103 Lr [262]	rutherfordium 104 Rf [261]	dubnium 105 Db [262]	seaborgium 106 Sg [266]	bohrium 107 Bh [264]	hassium 108 Hs [269]	meitnerium 109 Mt [268]	ununnilium 110 Uun [271]	unununium 111 Uuu [272]	ununbium 112 Uub [277]		ununquadium 114 Uuq [289]				

*Lanthanide series	lanthanum 57 La 138.91	cerium 58 Ce 140.12	praseodymium 59 Pr 140.91	neodymium 60 Nd 144.24	promethium 61 Pm [145]	samarium 62 Sm 150.36	europium 63 Eu 151.96	gadolinium 64 Gd 157.25	terbium 65 Tb 158.93	dysprosium 66 Dy 162.50	holmium 67 Ho 164.93	erbium 68 Er 167.26	thulium 69 Tm 168.93	ytterbium 70 Yb 173.04
* *Actinide series	actinium 89 Ac [227]	thorium 90 Th 232.04	protactinium 91 Pa 231.04	uranium 92 U 238.03	neptunium 93 Np [237]	plutonium 94 Pu [244]	americium 95 Am [243]	curium 96 Cm [247]	berkelium 97 Bk [247]	californium 98 Cf [251]	einsteinium 99 Es [252]	fermium 100 Fm [257]	mendelevium 101 Md [258]	nobelium 102 No [259]

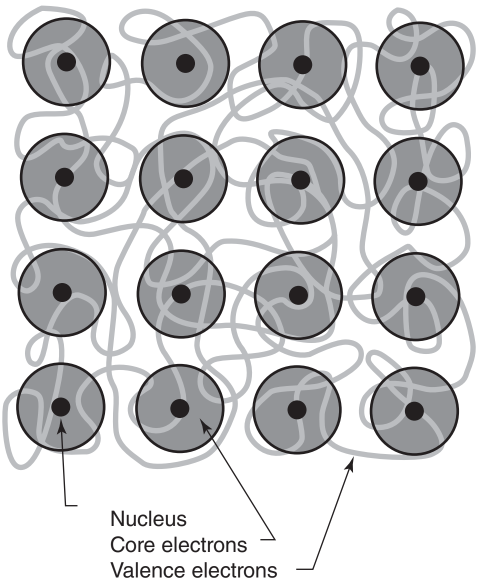
Figure 6.1 Periodic table.

Treating core and valence states differently is clearly a good idea. Although the Kohn–Sham equation is much simpler than the Hartree–Fock description of matter, the problem remains difficult. The disparate length and energy scales between the valence and core states make it difficult to use simple functions to describe the wave functions. Also, the number of electrons per atom in the periodic table varies by two orders of magnitude. Suppose we want to describe a molecule with both C and Pb atoms. Although the outer electronic configuration is the same, Pb contains 82 electrons and C only 6. A typical solution of the Kohn–Sham equation scales roughly as the cube of the number of electrons. As such, one might expect a Pb atom to be $(82/6)^3 \approx 2500$ times more difficult to solve than a C atom and that is just from the electron count. We have not considered the extra difficulties associated with the different energy and length scales.

This scaling factor can place a notable computational barrier on the types of systems that can be studied using quantum mechanics. Properties of small molecules with light elements are decidedly easier to compute than large molecules with heavy elements.

One of the first workers to address this problem was Hellmann, who recognized that without some approximations the chemical bond might never be understood by direct calculations [1]. Hellmann suggested treating the electronic states in a manner similar to the periodic table. He wrote the total energy of an atom as two terms. The first term focused on the core states. He argued that one could take the Thomas–Fermi kinetic energy of the core states and add to this the potential energy within the core region of the atom. This combination would result in a weak potential, which the valence electrons would experience. This idea of an effective core potential for the valence electrons was somewhat *ad hoc*, but Hellmann's ideas were to be justified by a number of workers. A picture of matter according to Hellmann's ideas is shown in Figure 6.2 where valence electrons interact to form bonds and core electrons remain inert.

Figure 6.2 Pseudopotential model. The outer electrons (valence electrons) move in a fixed arrangement of chemically inert ion cores. The ion cores are composed of the nucleus and core electrons.



6.1.1 Phillips–Kleinman Cancellation Theorem

In the late 1950's, James Phillips and Leonard Kleinman observed that the strong ion core potential could be replaced by a weak pseudopotential [2]. Morrel Cohen and Volker Heine cast the Phillips-Kleinman findings as a cancellation theorem [3].

We outline this theorem by considering a single atom with one valence electron, sodium being the prime example. We write an approximation for the valence state, ϕ_{v}:

$$\psi_{\mathrm{v}}(r) = \phi_{\mathrm{p}}(r) + \sum_{c} b_c \phi_c(r) \tag{6.1}$$

where ϕ_{p} is the pseudopotential wave function for the valence state and ϕ_c is a core state for the atom.

The more physics we can build into this wave function, the more accurate we expect it to be. We know several things about this wave function: (i) Away from the core region where the potential varies slowly, the wave function will also be slowly varying and easily represented by a simple plane wave or perhaps a Gaussian function. (ii) Near the nucleus of the atom ϕ_{v} will be "core like" and include core-like components. Incorporating ϕ_c in the wave function should satisfy this expectation. (iii) Valence states should be orthogonal to the core states.

The orthogonality condition is easy to show. We consider two wave functions, $(\phi_i,\ \phi_j)$, which are solutions to the Schrödinger equation: $\mathscr{H}\,\phi_i = E_i\phi_i$ and $\mathscr{H}\,\phi_j = E_j\phi_j$. If this is true, then the following must also be the case:

$$\begin{aligned}
&\int \phi_j^* H \phi_i \mathrm{d}^3 r = E_i \int \phi_j^* \phi_i \mathrm{d}^3 r \\
&\int \phi_j^* H \phi_i \mathrm{d}^3 r = \int (H\phi_j^*) \phi_i \mathrm{d}^3 r = E_j \int \phi_j^* \phi_i \mathrm{d}^3 r
\end{aligned} \tag{6.2}$$

We have used a trick in the second equation. Namely, we used $\mathscr{H}$ to operate on the wave function on the left of this operator. This is a valid operation provided the wave functions vanish at infinity, which they clearly do for a single atom. We also assume the eigenvalues are real, which they must be for a physical solution.

If all this is correct, we have

$$E_i \int \phi_j^* \, \phi_j \, \mathrm{d}^3 r = E_j \int \phi_j^* \, \phi_i \, \mathrm{d}^3 r \tag{6.3}$$

or

$$(E_i - E_j) \int \phi_j^* \, \phi_i \, \mathrm{d}^3 r = 0 \tag{6.4}$$

Assuming $E_i \neq E_j$, it must follow that

$$\int \phi_j^* \, \phi_i \, \mathrm{d}^3 r = 0 \tag{6.5}$$

i.e. the states are orthogonal. Clearly, this will be the case for valence and core states, as $E_\mathrm{v} \neq E_c$.

The sum in Eq. (6.1) includes all core states in the atom. In the case of sodium, this sum includes the 1s, 2s, 2p states. The coefficients, $\mathrm{b_t}$, are chosen so that the valence state is orthogonal to the core states, i.e. we have the following condition:

$$\int \phi_c^* \, \psi_\mathrm{v} \, \mathrm{d}^3 r = 0 \tag{6.6}$$

The orthogonality condition for the valence to core states yields

$$\int \phi_{c'}^*(r) \psi_\mathrm{v}(r) \, \mathrm{d}^3 r = \int \phi_{c'}^*(r) \phi_\mathrm{p}(r) \mathrm{d}^3 r + \sum_c b_c \int \phi_{c'}^*(r) \phi_c(r) \, \mathrm{d}^3 r \tag{6.7}$$

This determines b_c:

$$b_c = - \int \phi_{c'}^*(r) \phi_\mathrm{p}(r) \mathrm{d}^3 r = -\langle c | p \rangle \tag{6.8}$$

where we will use the compact notation $\langle c|p\rangle$ for the integral. Using this value for b_c, we can use the orthogonalized wave function to write

$$\begin{gathered} \mathscr{H} \psi_\mathrm{v} = E_\mathrm{v} \psi_\mathrm{v} \\ \mathscr{H} [\phi_\mathrm{p}(r) - \sum_c \langle c|p\rangle \, \phi_c(r)] = E_\mathrm{v} [\phi_\mathrm{p}(r) - \sum_c \langle c|p\rangle \, \phi_c(r)] \end{gathered} \tag{6.9}$$

We note that $\mathscr{H} \phi_c = E_c \phi$ and rearrange the terms:

$$\mathscr{H} [\phi_\mathrm{p}(r) + \sum_c (E_\mathrm{v} - E_c) \langle c|p\rangle \, \phi_c(r)] = E_\mathrm{v} \phi_\mathrm{p}(r) \tag{6.10}$$

We define a pseudopotential Hamiltonian, $\mathscr{H}_\mathrm{p}$, such that

$$\begin{aligned} \mathscr{H}_\mathrm{p} &= \mathscr{H} + \mathscr{V}_\mathrm{R} \\ \mathscr{V}_\mathrm{R} \, \phi_\mathrm{p} &= \sum_c (E_\mathrm{v} - E_c) \langle c|p\rangle \, \phi_c(r) \end{aligned} \tag{6.11}$$

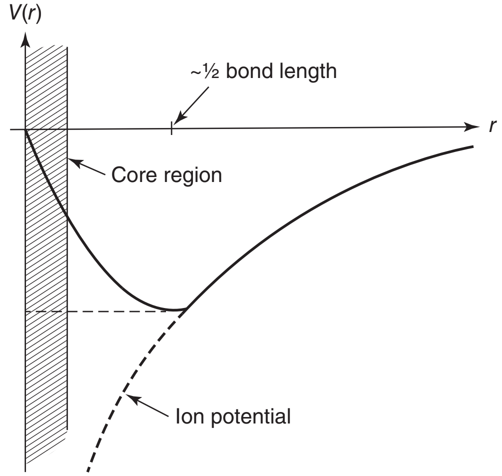


Figure 6.3 Schematic pseudopotential in real space.

and a pseudopotential, V_{p},

$$V_{\mathrm{p}}\phi_{\mathrm{p}} = (V + \mathscr{V}_{\mathrm{R}})\phi_{\mathrm{p}} \tag{6.12}$$

where V is the "all-electron" or "real" potential. The repulsive potential, $\mathscr{V}_{\mathrm{R}}$, cancels the strong Coulombic potential within the core region and results in a "weak" pseudopotential. This is illustrated schematically in Figure 6.3.

We are left with a "harmless" looking equation:

$$\mathscr{H}_{\mathrm{p}}\phi_{\mathrm{p}} = E_{\mathrm{v}}\phi_{\mathrm{p}} \tag{6.13}$$

In order to extract the pseudo wave function, we solve a Schrödinger equation that has the correct valence energy. What has changed is the Hamiltonian, which is now a pseudo-Hamiltonian containing not the real potential, but a pseudopotential. This pseudopotential, unlike Eq. (6.13), looks really obnoxious for several reasons. The chief problem is the $\mathscr{V}_{\mathrm{R}}$ part of the potential. This term is *not* a simple potential. Rather, it is a "nonlocal" operator that acts on the pseudo wave function and replaces it with a sum over the core states. Each core state is weighted by factor of $\langle c|p\rangle$, which is called a projection of the core state on the pseudo wave function. Even worse, the operator $\mathscr{V}_{\mathrm{R}}$ depends on the valence energy, E_{v}. This implies that we must solve a problem with an energy-dependent Hamiltonian. It does not get much worse from a computational perspective.

To summarize, Phillips and Kleinman transformed the all-electron problem to a pseudopotential problem, which only includes the valence electrons. In the all-electron problem, the Hamiltonian contains a strong potential that binds all the electrons; however, this potential is a simple function of position, which is good. But the bad features of the all-electron electron potential can be notable. The energy scale for the eigenvalues of the all-electron Hamiltonian can span orders of magnitude, and the wave functions must accurately describe the disparate length scales of the core and valence states. In the pseudopotential problem, the pseudo Hamiltonian contains a complicated nonlocal, energy-dependent operator, which is the difficult part. However, the pseudopotential is weak and the pseudo wave functions are simple. Roughly speaking, we have transformed the problem where the difficulty is focused on getting an accurate wave function to one where we now worry about getting an accurate pseudopotential.

In practice, no one uses (or almost no one uses) the pseudopotential as cast by Phillips and Kleinman. The chief accomplishment of their work was to show a prescription for separating the core and valence states, and to gain a better understanding of the electronic structure problem. For example, their work helps one understand why carbon and silicon are so different. Carbon lacks a p-state in the core. Therefore, there is no orthogonality condition on the 2p valence state. This is not the case for silicon where there exists a p-state in the core regime. In carbon, there is no $\mathscr{V}_\mathrm{R}$ term that results in a weak pseudopotential for the p valence states. There is such a potential for silicon. As such, the s and p valence states in carbon behave differently than those in silicon.

The pseudopotential theory has direct consequences in the chemistry of carbon versus that of silicon. Carbon exists in a variety of elemental structures ranging from graphene to graphite to diamond to fullerenes. The multiple structures of carbon reflect the ease of mixing s and p states, which can be traced to the energy required to promote an electron from the 2s-state to the 2p-state. Since the carbon 2p-states see no repulsive term from an orthogonality condition, the 2p-states are closer in energy to the 2s-state than is the case for the silicon 3s and 3p-states. Consequently, it is more difficult to form double and triple bonds in silicon when compared to carbon, e.g. carbon readily forms acetylene H—C≡C—H whereas H—Si≡Si—H (disilyne) is known to be unstable. One could argue that the Phillips–Kleinman cancellation theorem clarified the difference between biology (carbon chemistry) and geology (silicon chemistry). This is not a small feat for a relatively simple theorem.

6.2 Pseudopotentials Within Density Functional Theory

Although the Phillips–Kleinman cancellation theorem did not offer us a workable pseudopotential, it set the stage for later work. Contemporary approaches to the electronic structure problem are based on pseudopotentials constructed from the Kohn–Sham equation [4]. In this construction, one starts from the electronic structure of an atom. This is an easy problem to solve using the Kohn–Sham equation. The atom is taken to be spherically symmetric and the problem reduces to solving a one-dimensional equation. It is possible to solve the Kohn–Sham equation for a heavy element such as lead in a matter of seconds on a laptop computer. Molecules and clusters are another matter owing to the loss of spherical symmetry. For such systems, we definitely need pseudopotentials.

Consider a sodium atom for the purpose of designing a pseudopotential within the density functional theory. We know the eigenvalue, E_{3s}, and the corresponding wave function, $\psi_{3s}(r)$, for the valence electron from an atomic computation. We use this information as the starting point for the pseudopotential construction.

Good pseudopotentials must meet the following requirements: (i) The pseudopotential should bind only the valence electron: the 3s-electron for the case of the sodium atom. (ii) The eigenvalue of the corresponding valence electron should be identical to the full potential eigenvalue. (The full potential is also called the *all-electron* potential.) (iii) The pseudo wave function should be nodeless and *identical* to the "all-electron" wave function outside the core region.

For example, when a sodium pseudo wave function is constructed, $\phi_{3s}(r)$, it should have the following properties. It should be equal to the all-electron wave function, $\psi_{3s}(r)$, away from the core region. We can write this condition as $\phi_{3s}(r) = \psi_{3s}(r)$ for $r > r_c$ where r_c defines the size spanned by the *ion core*, i.e. the nucleus and core electrons. Typically, r_c is taken to be less than the distance corresponding to the maximum of the valence wave function, but greater than the distance of the outermost node.

The all electron wave functions and the pseudopotential wave functions for a Na atom are given in Figure 6.4.

In choosing the pseudo wave function within the core region, the integral of the pseudo charge density, i.e. square of the wave function, $|\phi_p(r)|^2$, within the core should be equal to the integral of the all-electron charge density. Without this condition, the pseudo wave function can be severely flawed. It can differ by a scaling factor from the all-electron wave function, that is, $\phi_p(r) = C \times \psi_{AE}(r)$ for $r > r_c$, where the constant, C, may differ from unity. This can be seen from the following:

$$1 = \int_0^\infty |\psi_{AE}(r)|^2 \, d^3r = \int_0^{r_c} |\psi_{AE}(r)|^2 \, d^3r + \int_{r_c}^\infty |\psi_{AE}(r)|^2 \, d^3r \qquad (6.14)$$

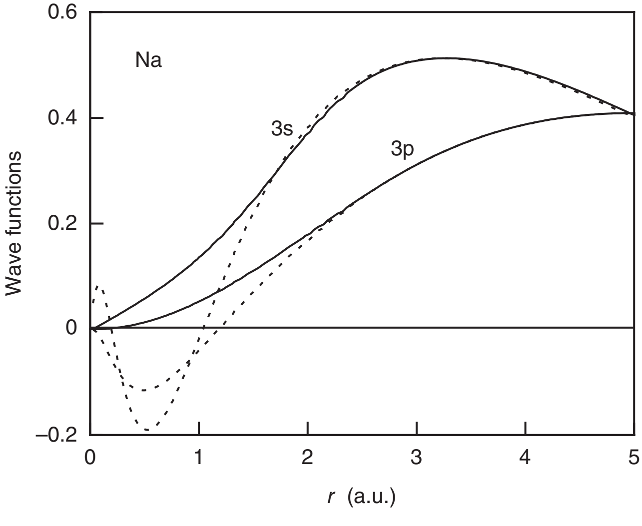


Figure 6.4 Pseudopotential wave functions compared to all-electron wave functions for the sodium atom. The all-electron wave functions are indicated by the dashed lines.

We can also write

$$1 = \int_0^{\infty} |\phi_{\mathrm{p}}(r)|^2 \, \mathrm{d}^3r = \int_0^{r_c} |\phi_{\mathrm{p}}(r)|^2 \, \mathrm{d}^3r + \int_{r_c}^{\infty} |\phi_{\mathrm{p}}(r)|^2 \, \mathrm{d}^3r \tag{6.15}$$

or if we take $\phi_{\mathrm{p}}(r) = C \times \psi_{\mathrm{AE}}(r)$,

$$1 = \int_0^{\infty} |\phi_{\mathrm{p}}(r)|^2 \, \mathrm{d}^3r = \int_0^{r_c} |\phi_{\mathrm{p}}(r)|^2 \, \mathrm{d}^3r + C \int_{r_c}^{\infty} |\psi_{\mathrm{AE}}(r)|^2 \, \mathrm{d}^3r \tag{6.16}$$

Combining Eqs. (6.14) and (6.16), we can write

$$\int_0^{r_c} |\phi_{\mathrm{p}}(r)|^2 \, \mathrm{d}^3r - \int_0^{r_c} |\psi_{\mathrm{AE}}(r)|^2 \, \mathrm{d}^3r = (1 - C) \int_{r_c}^{\infty} |\psi_{\mathrm{AE}}(r)|^2 \, \mathrm{d}^3r \tag{6.17}$$

$C \neq 1$ is bad. This will not occur if the left-hand side of Eq. (6.17) vanishes. This puts a constraint on the pseudo wave functions:

$$\int_0^{r_c} |\phi_{\mathrm{p}}(r)|^2 \, \mathrm{d}^3r = \int_0^{r_c} |\psi_{\mathrm{AE}}(r)|^2 \, \mathrm{d}^3r \tag{6.18}$$

The condition, $C = 1$, guarantees that any overlap of the pseudo wave function will be the same as for the all-electron wave functions. Pseudopotentials that reflect $C = 1$ are called "norm conserving." This nomenclature is a bit odd in that wave functions are always normalized, but here it means something different, i.e. the normalized pseudo wave function and the normalized all-electron wave functions are identical outside the core region.

There are many recipes for constructing pseudopotentials. We will focus on methods that are easy to understand and implement. In 1980, Kerker [5] proposed a straightforward method for constructing local density pseudopotentials that retained the norm-conserving criterion. He suggested that the pseudo wave function has the following form:

$$\phi_{\mathrm{p}}(r) = r^l \exp(p(r)) \quad \text{for } r < r_c \tag{6.19}$$

where $p(r)$ is a simple polynomial: $p(r) = -a_0 r^4 - a_1 r^3 - a_2 r^2 - a_3$ and

$$\phi_{\mathrm{p}}(r) = \psi_{\mathrm{AE}}(r) \quad \text{for } r > r_c \tag{6.20}$$

This form of the pseudo wave function for ϕ_{p} assures us that the function will be nodeless and have the correct behavior at large r. Kerker proposed the following criteria for fixing the parameters (a_0, a_1, a_2, and a_3):

1. The pseudo wave function should be norm conserving.
2. The all-electron and pseudo wave functions should have the same valence eigenvalue.
3. The pseudopotential wave function should be nodeless and be identical to the all-electron wave function for $r > r_c$.
4. The pseudo wave function must be continuous as well as the first and second derivatives of the wave function at r_c.

As indicated, there is some flexibility in constructing pseudopotentials within the core region ($r < r_c$). The non-uniqueness of the pseudo wave function was recognized early in its inception. This attribute can be exploited to optimize the convergence of the pseudopotentials for the basis of interest. Much effort has been made to construct pseudopotentials that require a minimum number of basis set functions, e.g. plane waves, to achieve a converged solution. Typically, shallow or "soft" pseudopotentials are characterized by a large core radius. As the core size is increased, the convergence between the all-electron and pseudo wave functions is postponed to larger distances; the quality of the pseudo wave functions will deteriorate and the the transferability of the pseudopotential between the atom and molecular environments becomes limited.

One straightforward approach to optimizing a pseudopotential is to build additional constraints into the polynomial given in Eq. (6.19). For example, suppose we write

$$p(r) = c_o + \sum_{n=1}^{N} c_n r^n \tag{6.21}$$

In Kerker's scheme, $N = 4$. However, there is no compelling reason for demanding that the series terminate at this particular point. If we extend the expansion, we may impose additional constraints. For example, we might try to constrain the reciprocal space expansion of the pseudo wave function so that beyond some momentum cutoff the function vanishes.

Troullier and Martins developed a very popular scheme by adding some physically motived constraints [6]. They wrote Eq. (6.21) as

$$p(r) = c_o + \sum_{n=1}^{6} c_{2n} r^{2n} \tag{6.22}$$

As usual, they constrained the coefficients to be norm conserving. In addition, they demanded continuity of the pseudo wave functions and the first four derivatives at r_c. They added a new type of constraint by demanding zero curvature of the pseudopotential at the origin. These potentials tend to be quite smooth and converge rapidly in reciprocal space.

Once the pseudo wave function is defined as in Eqs. (6.19) and (6.20) we can "invert" the Kohn–Sham equation and solve for the ion core pseudopotential, $V_{\text{ion,p}}$:

$$V^n_{\text{ion,p}}(\vec{r}) = E_n - V_{\text{H}}(\vec{r}) - V_{\text{xc}}[\vec{r}, \rho(\vec{r})] + \frac{\hbar^2 \nabla^2 \phi_{\text{p},n}}{2m\phi_{\text{p},n}} \tag{6.23}$$

This potential, when self-consistently screened by the pseudo charge density

$$\rho(\vec{r}) = -e \sum_{n,\text{occup}} |\phi_{\text{p},n}(\vec{r})|^2 \tag{6.24}$$

will yield an eigenvalue of E_n and a pseudo wave function $\phi_{\text{p},n}$. By construction, the pseudo wave function will agree with the all-electron wave function away from the core.

There are some important issues to consider about the details of this recipe. First, the potential is *state* dependent as written in Eq. (6.23), i.e. the pseudopotential is dependent on the quantum state n. This issue can be handled by recognizing the nonlocality of the pseudopotential, *i.e.*, the potential is not a simple function. The potential is different for an s-, p-, or d-electron. The nonlocality appears in the angular dependence of the potential, but not in the radial coordinate. Sometimes the pseudopotential is called "semi-local" instead of "nonlocal."

A related issue is whether the potential is highly dependent on the state energy, e.g. if the potential is fixed to replicate the 3s state in Na, will it also do well for the higher s states, i.e. 4s, 5s, 6s, …? Of course, one could also question how dependent the pseudopotential is on the atomic state used for its construction. For example, would a Na potential be very different for a $3s^13p^0$ versus a $3s^{1/2}3p^{1/2}$ configuration or a $3s^03p^1$ configuration? Finally, how important are "loosely bound" core states in defining the potential? For example, can one treat the 3d states in copper as part of the core and get good results? Or, do we need to consider it part of the valence shell?

Each of these issues has been carefully addressed in the literature [4]. We consider the last point first. In many cases, the separation between the core states and the valence states is clear. For example, in Si there is no issue that the core is composed of the $1s^22s^2p^6$ states. However, the core in Cu could be considered to be the $1s^22s^2p^63s^23p^63d^{10}$ configuration with the valence shell consisting of the $4s^1$ state. Or, one could consider the core to be the $1s^22s^2p^63s^23p^6$ configuration with the valence shell composed of the $3d^{10}4s^1$ states. On computational grounds, it would be much, much better to consider the correct Cu configuration as $1s^22s^2p^63s^23p^63d^{10}$. The number of valence states is just 1 instead of 11. Later, we will show that computationally the scaling of this problem goes as the cube of the number of states. So, carrying the d-states could result in an order of magnitude increase in the computational load. We are not that lucky. We must carry the d states as Cu is not at all like K, e.g. Cu does not burst into flames when dropped into water – K does. It is the outer 3d shell that distinguishes Cu from K.

Such issues are traditionally considered on a case by case basis. Fortunately, we can assess the matter in a straightforward manner. We can construct different pseudopotentials for all possible core–valence dichotomies and solve the Kohn–Sham equation. In the case of Cu, we would find that a pseudopotential constructed with the $3d^{10}$ state in the core yields poor results compared to an all-electron potential.

A related aspect of deciding how best to construct the core configuration is the issue of "core–valence" exchange-correlation interactions. In the all-electron exchange-correlation potential, the charge density is composed of the core and valence states; in the pseudopotential treatment only the valence electrons are included. This separation neglects terms that may arise between the overlap of the valence and core states.

There are well-defined procedures for including these overlap terms. It is possible to include a fixed charge density from the core and allow the valence overlap to be explicitly included. This procedure is referred to as a *partial core correction* [7]. This correction is especially important for divalent elements such as Zn, Cd, and Hg where the outermost filled d-shell can make some contribution to chemical bonding. But, in contrast to Cu, the contributions are not large and can be treated in an approximate manner without loss of accuracy. The importance of this correction can be tested by performing calculations with and without the partial core.

One might argue that the most accurate approach would be to include any "loosely bound" core states as valence states. This approach is often not computationally attractive in that it would increase the number of valence states, but there are other reasons. For example, the Zn core without the 3d states results in an ion core that must bind a dozen electrons, not just two electrons, i.e. the ion core pseudopotential corresponds to Zn^{+12}. A pseudopotential binding a dozen valence electrons is strong and deep compared to one binding only two valence electrons. We want pseudopotentials to be "weak and shallow," if possible.

In contrast, problems associated with the state dependence of the pseudopotential construction can be overcome with little computational effort. If the core electrons are tightly bound, the ion core potential is highly localized and often not sensitive to the ground state configuration used to compute the pseudopotential. There are well-defined tests for assessing the accuracy of the pseudopotential, especially in terms of the phase shifts. Also, it should be noted that higher excited states sample the tail of the pseudopotential in a region where there is no difference between the all-electron potential and the pseudopotential. As an example, the Na pseudopotential is usually converged to the all-electron potential by ~ 3 a.u. The 4s state of Na has a maximum near 10 a.u. and effectively samples only an all-electron-type potential. The details of the 4s state is largely unaffected by the details of a pseudopotential constructed by 3s state properties.

We need to make note of a problem with the local density approximation that is often conflated with the pseudopotential approximation. Consider moving an electron from the atom off to infinity. Classically, there is net residual charge on the atom of $+\text{e}$ when the electron is removed. We should have an attractive potential energy of $-\text{e}^2/r$ between the atom and the electron at a large separation. That does not happen in the local density approximation.

The local density approximation in its simplest form scales as $\rho^{1/3}$. This form does not yield a Coulomb-like interaction at large distances. Rather, the exchange-correlation potential will scale exponentially at large distances. This flaw comes from using an approximate density functional theory and not the pseudopotential. This issue appears when we deal with localized systems or with surfaces. While the electronic potential can be incorrect away from an atom, the total energy of the system does not suffer as much as one might think. Most of the electronic energy occurs from short range interactions where the local density approximation works better.

The chief complexity of pseudopotentials centers on nonlocality, which is often treated in Fourier space, but may also be expressed in real space. The interactions between valence electrons and pseudo-ionic cores may be separated into a local potential and a Kleinman and Bylander form [8] in *real space*:

$$V_{\text{ion}}^{p}(\vec{r})\phi_n(\vec{r}) = \sum_{a} V_{\text{loc}}(|\vec{r}_a|)\phi_n(\vec{r}) + \sum_{a,\ n,lm} G_{n,lm}^{a} u_{lm}(\vec{r}_a)\Delta V_l(r_a), \tag{6.25}$$

$$G_{n,lm}^{a} = \frac{1}{< \Delta V_{lm}^{a} >} \int u_{lm}(\vec{r}_a)\Delta V_l(r_a)\psi_n(\vec{r})\text{d}^3r, \tag{6.26}$$

and $< \Delta V^a_{lm} >$ is the normalization factor,

$$< \Delta V^a_{lm} >= \int u_{lm}(\vec{r}_a)\Delta V_l(r_a)u_{lm}(\vec{r}_a)\mathrm{d}^3r \tag{6.27}$$

where $\vec{r}_a = \vec{r} - \vec{R}_a$, and the u_{lm} are the atomic pseudopotential wave functions of angular momentum quantum numbers (l, m) from which the l-dependent ionic pseudopotential, $V_l(r)$, is generated. $\Delta V_l(r) = V_l(r) - V_{\mathrm{loc}}(r)$ is the difference between the l component of the ionic pseudopotential and the local ionic potential.

As a specific example, in the case of Na, we might choose the local part of the potential to replicate only the $l = 0$ component as defined by the 3s state. The nonlocal parts of the potential would then contain only the $l = 1$ and $l = 2$ components. In many cases, particularly when dealing with electronic materials such as Si and GaAs, the angular momentum for higher components than $l > 2$ are not significant in the ground state. In these systems, one can treat the summation over $l = 0, 1, 2$ to be complete.

The choice of which angular component is chosen for the local part of the potential can be arbitrary. It is often convenient to choose the local potential to correspond to the highest l-component of interest. This simplifies the projection operators, which become more complex with higher angular momentum. Again, these issues can be tested by choosing different components for the local potential and assessing the resulting solutions.

In Figure 6.5, the ion core pseudopotential for Na is presented using the Troullier–Martins formalism for creating pseudopotentials. The nonlocality of the potential is evident from the existence of three pseudopotentials corresponding to the s-, p-, and d-states. This is a typical result for simple metal atoms. The depth of the pseudopotential increases with the angular momentum

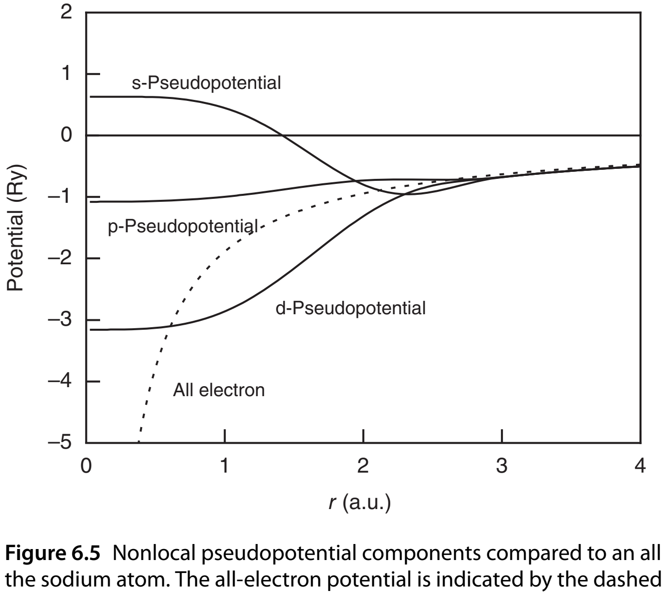


Figure 6.5 Nonlocal pseudopotential components compared to an all-electron potential for the sodium atom. The all-electron potential is indicated by the dashed lines.

component, i.e. the potential minimum for the d component is more attractive than for the s component.

References

1 Hellman, H. (1935). A new approximation method in the problem of many electrons. *Journal of Chemical Physics* 3: 61.
2 Phillips, J.C. and Kleinman, L. (1959). New method for calculating wave functions in crystals and molecules. *Physical Review* 116: 287.
3 Cohen, M.H. and Heine, V. (1961). Cancellation of kinetic and potential energy in atoms, molecules and solids. *Physical Review* 122: 1821.
4 Chelikowsky, J.R. and Cohen, M.L. (1992). Ab initio pseudopotentials for semiconductors. In: *Handbook on Semiconductors*, vol. 1 (ed. P. Landsberg), 59. Elsevier.
5 Kerker, G. (1980). Non-singular atomic pseduopotentials for solid state applications. *Journal of Physics C: Solid State Physics* 13: L189.
6 Troullier, N. and Martins, J.L. (1991). Efficient pseudopotentials for plane wave calculations. *Physical Review B* 43: 1993.
7 Louie, S.G., Froyen, S., and Cohen, M.L. (1982). Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Physical Review B* 26: 1738.
8 Kleinman, L. and Bylander, D.M. (1982). Efficacious form for model pseudopotentials. *Physical Review Letters* 48: 1425.

7

Methods for Atoms

> *A numerical goal without a method is nonsense.*
>
> — W. E. Deming

We have a formalism to describe the electronic structure of matter, but we have not outlined the methods to determine a numerical solution. For atoms, there are two methods often employed: one based on the variational method, the other based on direct numerical integrations.

7.1 The Variational Approach

We outlined in Chapter 3 a method for estimating the lowest energy state solution for a given Hamiltonian. If we have an approximate wave function, Ψ, we can estimate the energy of this state from

$$\hat{E} = \int \Psi^* \mathscr{H} \Psi \mathrm{d}^3 r \tag{7.1}$$

where we have assumed that the wave function is normalized. The *variational principle* states that $\hat{E} \geq E_0$, i.e. the approximate energy level is higher in energy (less negative) than the true energy level.

We might write a wave function with a few parameters, say $\Psi(\alpha, \beta, \gamma)$, and then find the energy, $E(\alpha, \beta, \gamma)$. By minimizing the energy with respect to these parameters, we can often get a good estimate of the energy.

7.1.1 Estimating the Energy of the Helium Atom.

We will illustrate this procedure, by returning to the simplest "many electron" atom – helium. Previously, we estimated the total energy by considering the hydrogen-like solution for a nucleus with a charge of $Z = 2$. This would be the solution for a helium atom if the electrons did not interact with each other. But, electrons do interact with each other. In order to account for the electron–electron interaction, we modify the wave function with this in mind. Again, the more we can build into a "trial" wave function based on physics, the better will be our estimate for the energy.

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

For example, we know that the presence of one electron can effectively "screen" the nucleus from the other electron. To account for this, we replace Z in the wave function by $Z = 2 - \sigma$. This is often done in textbook examples for helium [1]. If $\sigma = 0$, we expect no screening as the electron would see the full nuclear charge. At the other extreme, if the electron completely screened the charge, then we expect $\sigma = 1$.

We evaluate the total energy of the helium atom using the following approximate wave function:

$$\Psi = \sqrt{((2-\sigma)/\pi a_0)^3}\exp(-(2-\sigma)r/a_0) \tag{7.2}$$

These wave functions yield an estimate for the energy:

$$\begin{aligned}\hat{E}(\sigma) = \iint &((2-\sigma)^3/\pi a_0^3)\exp(-(2-\sigma)(r_1+r_2)/a_0)\\ &\times\left[-\frac{\hbar^2\nabla_1^2}{2m} - \frac{-\hbar^2\nabla_2^2}{2m} - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{|\vec{r}_1 - \vec{r}_2|}\right]\\ &\times((2-\sigma)^3/\pi a_0^3)\exp(-(2-\sigma)(r_1+r_2)/a_0)d^3r_1d^3r_2\end{aligned} \tag{7.3}$$

The resulting integrals are straightforward, albeit a bit tedious. The total electronic energy is given by

$$\hat{E}(\sigma) = -\frac{e^2}{a_0}(4 - (5/8 - 5/8\sigma - \sigma^2)) \tag{7.4}$$

We find the optimal value of σ by minimizing the total energy, i.e. we find $\hat{E}$ by setting $d\hat{E}/d\sigma = 0$. As expected, the optimal value of $\sigma = 5/16$ resides between the expected limits of 0 (no screening) and 1 (perfect screening.) This value of σ gives a total energy of

$$\hat{E}(\sigma) = -\frac{729}{256}\frac{e^2}{a_0} = -77.5\ \text{eV} \tag{7.5}$$

Recall that the correct value from experiment is $E(\text{He}) = E(\text{He}^+) - I = -54.4\ \text{eV} - 24.5\ \text{eV} = -78.9\ \text{eV}$; the error in the calculated value is our best estimate – less than a 2% error.

The value from our simple variational estimate of −77.5 eV is notably better than the previous estimate of −74.8 eV using hydrogen-like wave functions. Sophisticated trial wave functions have been employed and yield a value that is essentially in exact agreement with experiment. This finding not only reinforces the validity of the variational principle, but also validates quantum theory.

Example 7.1 ***Using Density Functional Theory to Find the Ionization Energy of a Helium Atom*** We can also illustrate the variational theorem using density functional theory. In the previous example, we examined the binding energy of an electron to helium atom. In doing so, we made no serious approximations to Schrödinger's equation; the answer is "exact," save some small corrections, e.g. we ignored the finite mass of the helium nucleus and maybe some relativistic issues. We can do this for a system such as helium, which has only two electrons. The issue is much worse when we deal with atoms with more than a few electrons.

In such a situation, it not only makes more sense to consider an approximate solution but it is also our only hope.

We mentioned several such approximations, Hartree–Fock and those based on density functional theory. Here, we show how these approximations can be used to estimate the ionization energy of a helium atom. Within density functional theory, we can write the total energy of the helium atom as

$$E = 2\int \psi^*(r)[-\frac{\hbar^2}{2m}\nabla^2 - \frac{2e^2}{r}]\psi(r)d^3r$$
$$+\frac{1}{2}\iint d^3rd^3r'\frac{\rho(r)\rho(r')}{|\vec{r}-\vec{r}'|} - \frac{3e^{2/3}}{4\pi}(3\pi^2)^{1/3}\int d^3r[\rho(r)]^{4/3} \tag{7.6}$$

The first term corresponds to a hydrogen-like atom with a nuclear charge of +2e; we have two electrons contributing to this term. The second term is the classical electrostatic term, or Hartree potential. The remaining term corresponds to the exchange energy. For simplicity, we ignore any attempt at including correlation. The charge density is given by

$$\rho(r) = -e\sum_{\text{occup}}|\psi(r)|^2 = -2e|\psi(r)|^2 \tag{7.7}$$

Again, there is a factor of two corresponding to one electron with spin up and one with spin down.

The simplest approach here is to assume a hydrogen-like atom again:

$$\psi(r) = \sqrt{(\beta^3/\pi)}\exp(-\beta r) \tag{7.8}$$

The prefactor of the exponential assures us that the wave function is normalized:

$$4\pi\int_0^\infty |\psi(r)|^2r^2dr = 4\beta^3\int_0^\infty \exp(-2\beta r)r^2dr = 1 \tag{7.9}$$

Inserting this trial wave function in the total energy expression and evaluating some integrals yields the following:

$$\hat{E}(\beta) = \frac{\hbar^2\beta^2}{m} - 4e^2\beta + 5e^2\beta/4 - \frac{81e^2\beta}{128}\left(\frac{6}{\pi^2}\right)^{1/3}\beta \tag{7.10}$$

We can rewrite this numerically as

$$\hat{E}(\beta) = \frac{\hbar^2\beta^2}{m} - 3.286\,e^2\beta \tag{7.11}$$

This yields a value of β as

$$\beta = \frac{1.643\,e^2m}{\hbar^2} = 1.643/a_0 \tag{7.12}$$

Recall from our variational approach for the "real" Hamiltonian that we had taken a similar expression for the wave function. We can identify $(2-\sigma)/a_0$ with β. From density functional theory, we predict a value of $\sigma \approx 2 - 1.643 = 0.357$ compared to $\sigma = 5/16 = 0.3125$. This is not too bad, but the total energies are a different matter. Density functional theory as we have implemented yields an estimate for the total energy of

$$\hat{E}(\beta) = e^2a_0\beta^2 - 3.286\,e^2\beta = -2.699\,e^2/a_0 = -73.4\text{ eV} \tag{7.13}$$

This value can be compared with our estimate using variational theory using the "real Hamiltonian," which yielded −77.5 eV and to the experimental value of −78.9 eV. In fact, our simple estimate with unaltered hydrogen-like wave functions was −74.8 eV.

This looks bad for density functional theory. However, the situation is not as bad as it might appear. As noted earlier, we typically have no interest in "absolute energies." We are interested in "relative energies." The situation is similar to thermodynamics. In thermodynamics, one is always interested in the *change* in the energy, or perhaps the free energy. As an example, suppose we are interested in calculating the binding energy of a molecule composed of elements A and B. We could find the energy of the molecule, $E(\text{AB})$, but by itself this would tell us nothing. We would rather know what the energy was compared to the individual constituents, $E(\text{A})$ and $E(\text{B})$. The difference would tell us whether A and B might react to form the molecule AB. What we need to do is to calculate the quantity $E(\text{AB}) - E(\text{A}) - E(\text{B})$. If we made an error in all these quantities, e.g. by a common factor of 5%, our net error would *not* be 5% of the absolute energy, but of the relative energies. Again, this is a crucial issue in applying approximate density functional theory.

We attempt the following: We will calculate the ionization of helium by calculating: $I = E(\text{He}^+) - E(\text{He})$. You might argue that we already know the He^+ energy as it is a hydrogen-like atom. We pretend not to know any better. We use the same expression as for $E(\text{He})$, save the obvious change–we remove an electron. This removal reduces the first two terms in Eq. (7.10), the kinetic and potentials energies, by a factor of two as they scale linearly with the number of electrons. The next term, the Hartree energy, is reduced by a factor four as the charge appears squared; the final term, the exchange energy, is reduced by a factor of $2^{4/3}$ as the charge appears to the 4/3 power. The resulting expression for the total energy of He^+ is

$$E_{\text{total}} = \beta^2 - 4\beta + 5\beta/4 - \frac{81}{128}\left(\frac{6}{\pi^2}\right)^{1/3}\beta \tag{7.14}$$

Again, we combine terms by evaluating the various factors and find

$$E_{\text{total}} = \beta^2/2 - 1.6866\beta \tag{7.15}$$

The lowest energy for He^+ occurs for

$$\beta_{\text{min}} = 1.900\ \text{a.u.} \quad E_{\text{min}} = -1.805\ \text{a.u..} = -49.10\ \text{eV} \tag{7.16}$$

The net ionization energy is now

$$I = E(\text{He}^+) - E(\text{He}) = (-49.10) - (-73.4)\ \text{eV} = 24.3\ \text{eV} \tag{7.17}$$

The experimental value is 24.6 eV. The error is $\sim 1\%$. This is better than using the variational principle with a hydrogen-like wave function.

Our result surely must be counterintuitive. The "exact" Hamiltonian yields a less accurate value with the same form of the basis. The secret to our success is clearly error cancellation. We know that the exact energy for a He^+ ion is −2.0 a.u.

or −54.4 eV. This is *not* in agreement with our simplified density functional theory, which yields an energy for a He^+ ion of −1.80 a.u. or −49.1 eV. Density functional theory makes an error of about 5 eV (too small) for both the ionized and neutral states of He.

Of course, we should remember that this example took the same, very simple basis for both the exact and approximate Hamiltonian. For the exact Hamiltonian, we know that our answer will approach the measured ionization energy as we continually improve the basis. It has to do so according to the variational theorem. For an approximate Hamiltonian, we have no such guarantee.

Maybe we were just lucky in getting such a good number for helium with our simple basis. We are not guaranteed that a better basis (one with more parameters to optimize) will work as well. We could test this issue by making sure our energies are converged to the lowest values for the helium atom and for the helium ion. If our converged energies differ from experiment, that would indicate a failure of the functional.

One could argue that we just need to improve the functional. Well and good to say this, but how to do it is an ongoing research problem.

7.2 Direct Integration

Example 7.2 ***The Hydrogen Atom*** Free atoms possess a property that considerably helps in solving for the electronic structure problem. We take an isolated atom to be spherically symmetric. In this case, the problem resembles that of a hydrogen atom. If the potential is spherical, i.e. $V(r) = V(|\vec{r}|)$, we solve a Schrödinger equation of the form similar to that of Eq. (2.20).

$$-\frac{\hbar^2}{2m}\frac{d^2u}{dr^2} + \frac{l(l+1)\hbar^2}{2mr^2}\,u + V(r)\,u = E\,u \tag{7.18}$$

We can integrate this one-dimensional equation directly. We outline some essential features of numerical methods based on Taylor series expansions. This is standard fare as outlined in textbooks on numerical methods [2].

We consider a first-order dimensional equation:

$$\frac{df}{dx} = g(f,x); \quad f(0) = f_0 \tag{7.19}$$

We need to have a boundary condition; we assume some value, f_0, at $x = 0$. Consider a function, $f(x)$, where we expand around a point x:

$$f(x+\delta x) = f(x) + \frac{df}{dx}\,\delta x + \cdots \tag{7.20}$$

We can make use of the differential equation, to write

$$f(x+\delta x) = f(x) + g(f,x)\,\delta x \tag{7.21}$$

Provided we take a sufficiently small value of δx, the equal sign is justified. We can step the function as follows:

$$f(\delta x) = f(0) + g(f(0),0)\,\delta x = f_0 + g(f_0,0)\,\delta x \tag{7.22}$$

We know all the terms on the right-hand side of this function. We know $f(0)$ from our boundary condition. We know $g(f_0, 0)$ from the given differential equation and we specify a value for δx to step the function. Of course, δx needs to be sufficiently small to make this numerical method work. We do not want $f(\delta x)$ to change much from $f(0)$ or our truncated Taylor series is not likely to be correct.

We can try to step this function again:

$$f(2\delta x) = f(\delta x) + g(f(\delta x), \delta x)\ \delta x \tag{7.23}$$

Again, we know all the terms on the right-hand side using our solution for $f(\delta x)$ so we now know $f(2\delta x)$. We can continue to execute this algorithm and find $f(x)$ as far as we are willing to step out the function. This procedure is known as the *explicit Euler method* [2].

The explicit Euler method works fine if δx is made sufficiently small, which can be an issue if the function $g(f, x)$ is difficult to evaluate. We can make a simple improvement on explicit Euler by recognizing that the key assumption in the Euler method is a linear step. Suppose we consider starting at the point $f(x + \delta x)$ and going backward to find $f(x)$:

$$f(x) = f(x + \delta x) - g(f(x + \delta x, x + \delta x)\ \delta x \tag{7.24}$$

We can rewrite this as

$$f(x + \delta x) = f(x) + g(f(x + \delta x), x + \delta x)\ \delta x \tag{7.25}$$

This looks very much like Eq. (7.21), save that the value of the function g is now evaluated at $f(x + \delta x)$ and is called the *implicit Euler method*. We combine Eqs. (7.21) and (7.25).

$$f(x + \delta x) = f(x) + \frac{1}{2}[g(f(\delta x), \delta x) + g(f(x + \delta x), x + \delta x)]\ \delta x \tag{7.26}$$

This expression is easy to understand. The value of the function at $x + \delta x$ is the average of the slope at x and at $x + \delta x$ times δx. This is a better estimate than just considering the slope at x. There remains a problem, which is a major drawback of the implicit Euler method. Namely, how does one calculate the value of the function $g(f(x + \delta x), x + \delta x)$ when $f(x + \delta x)$ is unknown?

We can implement an iterative procedure that works very well and makes a reasonable estimate for $f(x + \delta x)$. We proceed in two steps. First, we employ the explicit Euler step to estimate the value of $f(\delta x)$:

$$f_{\text{ee}}(\delta x) = f_0 + g(f_0, 0)\ \delta x \tag{7.27}$$

where f_{ee} is the explicit Euler estimate. We can use f_{ee} to improve on this estimate:

$$f(\delta x) = f_0 + \frac{1}{2}[g(f_0, 0) + g(f_{\text{ee}}(\delta x), \delta x)]\ \delta x \tag{7.28}$$

We can generalize this as follows for a point x in general.

$$\begin{aligned} f_{\text{ee}}(x + \delta x) &= f(x) + g(f(x), x)\ \delta x \\ f(x + \delta x) &= f(x) + \frac{1}{2}[g(f(x), x) + g(f_{\text{ee}}(x + \delta x), x + \delta x)]\ \delta x \end{aligned} \tag{7.29}$$

We use the values of the function $f(x)$ and explicit Euler to get an estimate of the function $f(x+\delta x)$ given by $f_{\text{ee}}(x+\delta x)$, which is then used to find an average slope and obtain a better estimate of $f(x+\delta x)$. This procedure is called the "Heun method" or the second-order "Runge-Kutta method" [2].

While we now have an iterative method for solving a first-order differential equation, the problem that we want to solve is *second order*. We can still use the method. For simplicity, let us take $l = 0$ in Eq. (7.18) and use atomic units ($e = \hbar = m = 1$) along with the potential for hydrogen:

$$\frac{\mathrm{d}^2 u}{\mathrm{d}r^2} = 2\left(-\frac{1}{r} - E\right)\ u \tag{7.30}$$

We define a new variable as $v = \mathrm{d}u/\mathrm{d}r$ and we now have a set of coupled first-order differential equations to solve instead of one second-order equation:

$$\begin{aligned}
\frac{\mathrm{d}u}{\mathrm{d}r} &= v \\
\frac{\mathrm{d}v}{\mathrm{d}r} &= 2(V - E)\ u = -2\left(\frac{1}{r} + E\right)\ u
\end{aligned} \tag{7.31}$$

We expand the functions around $r = 0$. It is easy to show that within this limit

$$\begin{aligned}
v &= 1 - 2r + \cdots \\
u &= r - r^2 + \cdots
\end{aligned} \tag{7.32}$$

This gives us initial conditions for integrating the Schrödinger equation: $u(0) = 0$ and $v(0) = 1$.

There remain some issues here. For example, we can use the Heun method to integrate the set of radial equations for (u, v), but we have to be careful near $r = 0$ owing to the divergence of the Coulomb potential. Also, what do we use for δx (or better δr) in Eq. (7.29)? For that matter, what value do we use for E?

The issue about the singularity can be handled by not starting exactly at $r = 0$; after all, we know from the boundary conditions how the functions behave at the origin (Eq. (7.32)). We can try a large value of δr and keep decreasing it until the solutions converge. A value of $\delta r = 0.001$ a.u. works well. This is a small value, but using a computer we can do many, many steps in a very, very short time.

We need to consider a value for E. We use the limiting behavior of the wave function. We know when $r \to \infty$ that we must have $R \to 0$. If this condition is not met, we cannot normalize the wave function.

This behavior is illustrated in Figure 7.1. We know that for the 1s state, the energy of the hydrogen atom is $E_{1s} = -0.5$ a.u. If we guess this value, the wave function is well behaved and goes to zero as r increases. Suppose we assume a value of $E_{1s} = -0.499$ a.u., it would seem to be a pretty good guess. However, it is not as good as one might think. The value of the radial wave function starts to diverge markedly from the exact solution by $r \approx 7$ a.u. A guess of $E_{1s} = -0.501$ a.u. is equally bad, save the divergence changes sign. By examining the value of the wave function for large values of r, it is possible to obtain a very accurate value for the energy levels in question. We know that the true energy level must reside between $E_{1s} = -0.499$ a.u. (negative divergence) and $E_{1s} = -0.501$ a.u. (positive divergence).

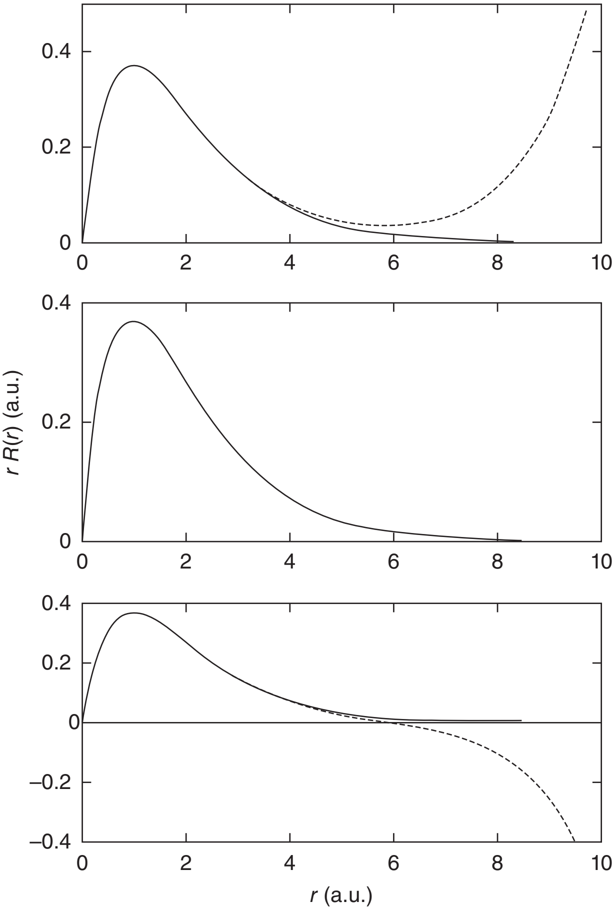


Figure 7.1 Integration of the Schrödinger equation for the hydrogen atom. The 1s state for r times the radial function $R(r)$ is shown. In the top panel a guess for the energy of the 1s state is taken to be $E_{1s} = -0.501$ a.u. In the middle panel the guess is $E_{1s} = -0.500$ a.u. In the bottom panel the guess is $E_{1s} = -0.499$ a.u. The exact solution for the 1s state is shown by a solid line. The integrated solutions are shown by a dashed line. For the correct guess of $E_{1s} = -0.500$ a.u. there is no difference discernible between the numerical integration and the exact solution.

We find other energy levels by making different guesses for the energy. For example, if we wanted the energy for the 2s level we can take a guess near $E_{2s}= -0.125$ a.u. This is shown in Figure 7.2 where again we take guesses near the exact value. Of course, this exercise supposes that we know what reasonable guesses make sense. This is not a severe limitation as the computer time to integrate the one-dimensional radial equation is almost "nothing." While we did these examples with $l = 0$, it is not much more difficult to consider high values of l.

The direct numerical integration works very well for systems with spherical symmetry. It is fast and easy to implement. It is the method of choice, even for atoms that have more than one electron.

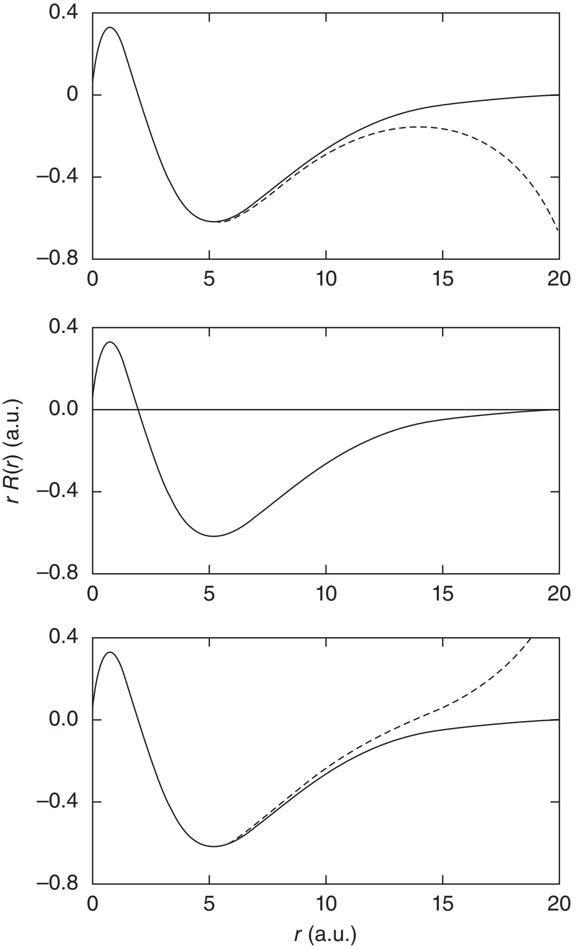


Figure 7.2 Integration of the Schrödinger equation for the hydrogen atom. The 2s state for *r* times the radial function *R*(*r*) is shown. In the top panel a guess for the energy of the 2s state is taken to be $E_{2s} = -0.126$ a.u. In the middle panel the guess is $E_{2s} = -0.125$ a.u. In the bottom panel the guess is $E_{2s} = -0.124$ a.u. The exact solution for the 2s state is shown by a solid line. The integrated solutions are shown by a dashed line. For the correct guess of $E_{1s} = -0.125$ a.u. there is no difference discernible between the numerical integration and the exact solution.

7.2.1 Many-electron Atoms Using Density Functional Theory

Calculations for many-electron atoms can be implemented within the framework of density functional theory. For an atom with an atomic number, Z, we need to solve the following equation:

$$\left(\frac{-\hbar^2\nabla^2}{2m} - \frac{-Ze^2}{r} + V_{\mathrm{H}}(\vec{r}) + V_{\mathrm{xc}}[n(\vec{r})]\right)\phi_i(\vec{r}) = E_i\phi_i(\vec{r}) \tag{7.33}$$

Since the potential between the nucleus and an electron is spherically symmetric, the equation is one dimensional in the radial coordinate. However, the problem is more complex owing to the many-electron terms. The Hartree and exchange-correlation potentials require knowledge of the electronic charge density as noted in Chapter 5. The Hartree potential, V_{H}, can be calculated from

$$\begin{aligned}\nabla^2 V_H(r) &= -4\pi e\rho(r)\\ \rho(r) &= e\sum_{i,\mathrm{occup}}|\phi(r)|^2\end{aligned} \tag{7.34}$$

where the charge density, ρ, is given by summing over the square of the occupied states. The exchange correlation potential, V_{xc}, is a functional of the charge density: $V_{xc} = V_{xc}[\rho]$. This presents an annoying issue. To find the density we need to know the wave functions from the Kohn–Sham equation, but this equation cannot be solved unless we know the potentials from the density, or unless we know the wave functions. This conundrum can be resolved by finding a *self-consistent* solution.

An initial guess is made for the density, sometimes using Thomas–Fermi theory as outlined in Chapter 5. Or, one can start from scratch by assuming a bare nucleus and solving for the 1s state, and then using its density as a starting point, solve for the 2s state and keep marching along until all the states are known.

Given that an "input" density allows one to solve for the Hartree and exchange-correlation potentials. Using these "input" potentials, the Kohn–Sham equation can be solved by direct integration using a methodology similar to that outlined in the previous section The resulting wave functions are then used to find a new "output" density. In the case of an atom, it is often feasible to use this "output" density to create new "input" potentials directly. However, often a combination of the "input" and "output" potentials is used. The process is continued until the input and output potentials (or densities) are the same within some specified tolerance. The self-consistent process is illustrated in Figure 7.3.

Codes for the electronic structure of atoms are fairly routine and very efficient. Even for the heavy atoms, e.g. a gold or a lead atom, the Kohn–Sham equations can be solved self-consistently within a few seconds on a laptop computer.

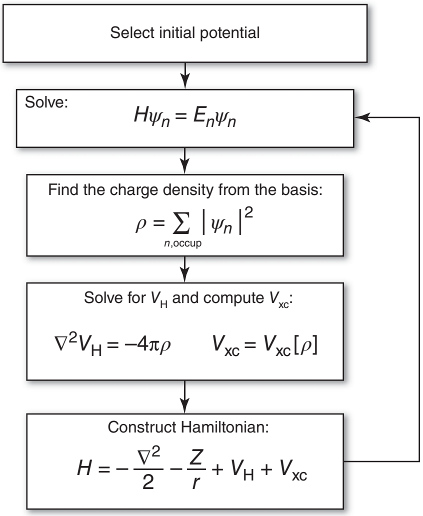


Figure 7.3 Schematic for creating a self-consistent field solution of the Kohn–Sham equation for an atom. Z is the atomic number of the atom. Atomic units ($e = m = \hbar = 1$) are used in the equations. The cycle is repeated until the input and output potentials (or charge density) are equal within some tolerance.

References

1 Merzbacher, E. (1999). *Quantum Mechanics*, 3e. Wiley.

2 Chapra, S.C. (2011). *Applied Numerical Methods W/MATLAB: For Engineers & Scientists*, 3e. McGraw-Hill.

8

Methods for Molecules, Clusters, and Nanocrystals

> *In so far as quantum mechanics is correct, chemical questions are problems in applied mathematics.*
>
> — Henry Eyring

Isolated atoms are special. If we assume space is isotropic, which is likely going to be the case save near blackholes, atoms are spherically symmetric. We can obtain a solution of the electronic structure problem for these systems by solving a one-dimensional problem. Even for heavy elements such as uranium or gold, where the length and energy scales between the tightly bound core states and valence states are quite disparate, a solution to the numerical problem can readily be had by employing a nonuniform grid.

Interacting atoms are a different story. The moment we start dealing with three-dimensional objects, the problem becomes much more difficult. The angular and radial coordinates can no longer be separated. The quantum numbers used to label atoms, (n, l, m, m_s), no longer work, save in an approximate sense.

8.1 The H_2 Molecule: Heitler–London Theory

Before we consider solutions for complex systems such as large molecules or clusters, we review one of the first approaches to the quantum theory of molecules. The theory is from Walter Heitler and Fritz London, who demonstrated that quantum methods predict the formation of a molecule when hydrogen atoms interact.

We write the Hamiltonian for a hydrogen molecule as follows:

$$\begin{aligned} H = &-\frac{\hbar^2\nabla_1^2}{2m} - \frac{\hbar^2\nabla_2^2}{2m} - \frac{e^2}{r_1} - \frac{e^2}{r_2} \\ &+\frac{e^2}{|\vec{r}_2 - \vec{r}_1 + \vec{R}|} - \frac{e^2}{|\vec{r}_1 - \vec{R}|} - \frac{e^2}{|\vec{r}_2 + \vec{R}|} + \frac{e^2}{R} \end{aligned} \tag{8.1}$$

where the geometry is as defined in Figure 8.1. The coordinates of the two electrons are given in terms of $(\vec{r}_1, \vec{r}_2)$. We take the distance between the two protons as $\vec{R}$. We use the Born–Oppenheimer approximation to separate the proton coordinates from the electron coordinates, and treat the protons as classical objects.

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

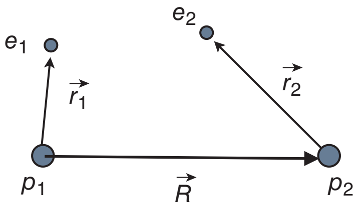


Figure 8.1 Coordinates for molecular hydrogen.

We also assume infinite masses for the protons, which is not a stretch as they are 1836 times more massive than an electron.

The first line in Eq. (8.1) for the Hamiltonian represents isolated hydrogen atoms labeled by (1,2). We consider this explicitly and write

$$H = H(H_1) + H(H_2) + \frac{e^2}{|\vec{r}_2 - \vec{r}_1 + \vec{R}|} - \frac{e^2}{|\vec{r}_1 - \vec{R}|} - \frac{e^2}{|\vec{r}_2 + \vec{R}|} + \frac{e^2}{R} \tag{8.2}$$

where $H(H_1), H(H_2)$ represent isolated H atoms that are labeled #1 and #2. The remaining terms can be considered a "perturbation" on the hydrogen atoms. The electron–electron repulsion term is the usual "difficult" term as it couples the coordinates for the two electrons. Otherwise, we could try a wave function with separable coordinates for the two hydrogen atoms.

Following Heitler and London [1], we construct a trial wave function to estimate the binding energy and the bond length of a hydrogen molecule. An obvious and natural choice is to take wave functions from the hydrogen atom. In the limit of weakly interacting H atoms, this will surely be correct. Such wave functions are solutions if we consider only the first two terms in Eq. (8.2). Of course, there is no reason to believe that such wave functions will be highly accurate when the atoms form a bond.

For notation purposes, we will define the following:

$$\begin{aligned} a(1) &= \frac{1}{\sqrt{\pi}} \exp(-r_1) \\ b(1) &= \frac{1}{\sqrt{\pi}} \exp(-|\vec{r}_1 + \vec{R}|) \\ a(2) &= \frac{1}{\sqrt{\pi}} \exp(-|\vec{r}_2 - \vec{R}|) \\ b(2) &= \frac{1}{\sqrt{\pi}} \exp(-r_2) \end{aligned} \tag{8.3}$$

where the distance is measured in atomic units (see the appendix on units). This basis has the following interpretation. The orbital for electron #1 can be localized on either proton, so we consider two orbitals, $a(1)$ and $b(1)$; the former is an atomic function localized on proton #1 and the latter is localized on proton #2. We use the same convention for electron #2. This is appropriate because we must allow for the possibility of an electron being localized on either proton. Heitler–London proposed the following wave function for the hydrogen molecule:

$$\psi(\vec{r}_1, \vec{r}_2) = A \times [a(1)b(2) + a(2)b(1)] \tag{8.4}$$

The constant A is the normalization factor. The wave function ψ allows us to consider the possibility of electron #1 being on proton #1 and electron #2 being on proton #2, or vice versa.

At first glance, this wave function appears to violate the Pauli principle. When we discussed Hartree–Fock, we insisted on writing a many-electron wave function so that it was antisymmetric: $\psi(\vec{r}_1, \vec{r}_2) = -\psi(\vec{r}_2, \vec{r}_1)$. In Eq. (8.4), this is not the case as

$$\psi(\vec{r}_2, \vec{r}_1) = A \times [a(2)b(1) + a(1)b(2)] = \psi(\vec{r}_1, \vec{r}_2) \tag{8.5}$$

Have we made a mistake? No, but this equation is incomplete. We left out spin. If we write down a wave function for the spin states, we can fix the problem:

$$\chi(1,2) = [\chi_1(\uparrow)\chi_2(\downarrow) - \chi_1(\downarrow)\chi_2(\uparrow)] \tag{8.6}$$

The spin states can be up (↑) or down (↓). Here we take a combination that puts the electrons in opposite spin states. Now the wave function behaves properly as

$$\chi(2,1) = [\chi_2(\uparrow)\chi_1(\downarrow) - \chi_2(\downarrow)\chi_1(\uparrow)] = -\chi(1,2) \tag{8.7}$$

We form a wave function with *both* spatial and spin coordinates: $\Psi(\vec{r}_1, \vec{r}_2) = \psi(\vec{r}_1, \vec{r}_2)\chi(1,2)$ where clearly $\Psi(\vec{r}_1, \vec{r}_2) = -\Psi(\vec{r}_2, \vec{r}_1)$. One could argue that a wave function could also be written with the spatial part being antisymmetric and the spin part being symmetric. The arbiter of the two choices is the total energy of the system. The lower energy state will correspond to the ground state. We shall see that an antisymmetric spatial part is not energetically favorable.

The energy of the electronic state can be estimated from

$$E_{H_2}(R) = \frac{\iint \psi^*(\vec{r}_1, \vec{r}_2)\mathrm{H}(\vec{r}_1, \vec{r}_2)\psi(\vec{r}_1, \vec{r}_2)\, \mathrm{d}^3r_1\, \mathrm{d}^3r_2}{\iint \psi^*(\vec{r}_1, \vec{r}_2)\psi(\vec{r}_1, \vec{r}_2)\, \mathrm{d}^3r_1\, \mathrm{d}^3r_2} \tag{8.8}$$

Since the Hamiltonian does not contain a spin dependence, the spin-dependent orbitals can be integrated out immediately. While the evaluation of the remaining spatial part of the integral is straightforward, it is surely tedious. We have normalized the orbitals to unity, so that the integral in the denominator can be expressed as

$$\begin{aligned} \iint \psi^*(\vec{r}_1, \vec{r}_2)\psi(\vec{r}_1, \vec{r}_2)\, \mathrm{d}^3r_1\, \mathrm{d}^3r_2 &= 2 + 2S^2 \\ S = \int a(1)b(1)\, \mathrm{d}^3r_1 &= \int a(2)b(2)\, \mathrm{d}^3r_2 \end{aligned} \tag{8.9}$$

The integral, S, is termed an "overlap" matrix as the hydrogen orbitals within the integral are localized on different sites. It can be written as an integral over the spherical coordinates (r, θ, ϕ):

$$S = \frac{1}{\pi}\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \sin(\theta)\mathrm{d}\theta \int_0^{\infty} r^2\mathrm{d}r \exp(-r)\ \exp\left(-\sqrt{r^2 + R^2 + 2rR\cos(\theta)}\right) \tag{8.10}$$

This has a solution:

$$S = \exp(-R)\,(1 + R + R^3/3) \tag{8.11}$$

We now have an expression for the wave function as

$$\psi(\vec{r}_2, \vec{r}_1) = \frac{[a(2)b(1) + a(1)b(2)]}{\sqrt{2(1 + S^2)}} \tag{8.12}$$

The total energy, E_{H_2}, can be expressed as the following integral:

$$\begin{aligned} E_{\mathrm{H}_2}(R) = & \frac{1}{2(1+S^2)} \iint (a(1)b(2) + a(2)b(1)) \left(\mathrm{H}_{\mathrm{H}_1} + \mathrm{H}_{\mathrm{H}_2} + \frac{1}{R} \right. \\ & \left. - \frac{1}{|\vec{r}_1 - \vec{R}|} - \frac{1}{|\vec{r}_2 + \vec{R}|} + \frac{1}{|\vec{r}_2 + \vec{R} - \vec{r}_1|} \right) (a(1)b(2) + a(2)b(1))\, \mathrm{d}^3 r_1 \mathrm{d}^3 r_2 \end{aligned} \tag{8.13}$$

Before we proceed to address this “messy” integral, we can test the limit $R \to \infty$. In this limit, $S \to 0$ and $a(2)b(1) \to 0$, i.e. the overlap between the orbitals vanishes. This leaves

$$E_{\mathrm{H}_2}(R \to \infty) = \iint (a(1)b(2)) \left(\mathrm{H}_{\mathrm{H}_1} + \mathrm{H}_{\mathrm{H}_2} \right) (a(1)b(2))\, \mathrm{d}^3 r_1 \mathrm{d}^3 r_2 \tag{8.14}$$

which can easily be rewritten as

$$\begin{aligned} E_{\mathrm{H}_2}(R \to \infty) &= \int a(1) \left(\mathrm{H}_{\mathrm{H}_1} \right) a(1) \mathrm{d}^3 r_1 + \int b(2) \left(\mathrm{H}_{\mathrm{H}_2} \right) b(2) \mathrm{d}^3 r_2 \\ E_{\mathrm{H}_2}(R \to \infty) &= E_{\mathrm{H}} + E_{\mathrm{H}} = -1 \text{ a.u.} \end{aligned} \tag{8.15}$$

The limit is correct. At infinite separation, we obtain the energy of two hydrogen atoms. This energy limit will serve as our reference energy. The integral for finite R is really unpleasant. The analytic solution can be written as

$$E_{\mathrm{H}_2}(R) = -1 + \frac{\mathrm{H}_0 + \mathrm{H}_1}{1 + S^2} \tag{8.16}$$

The first term reflects the energy of the isolated hydrogen atoms. H_0 and H_1 are given by

$$\begin{aligned} \mathrm{H}_0 &= \iint a^2(1) b^2(2) \left(\frac{1}{R} - \frac{1}{|\vec{r}_1 - \vec{R}|} - \frac{1}{|\vec{r}_2 + \vec{R}|} + \frac{1}{|\vec{r}_2 + \vec{R} - \vec{r}_1|} \right) \mathrm{d}^3 r_1 \mathrm{d}^3 r_2 \\ \mathrm{H}_1 &= \iint a(1)b(1)a(2)b(2) \left(\frac{1}{R} - \frac{1}{|\vec{r}_1 - \vec{R}|} - \frac{1}{|\vec{r}_2 + \vec{R}|} + \frac{1}{|\vec{r}_2 + \vec{R} - \vec{r}_1|} \right) \mathrm{d}^3 r_1 \mathrm{d}^3 r_2 \end{aligned} \tag{8.17}$$

H_0 is the Hartree or Coulomb integral as we can consider the charge density of electron to be given by $(a^2(1), b^2(2))$ for electrons (1,2), respectively. H_1 is the exchange integral. The Coulomb and exchange integrals can be evaluated in a straightforward manner to yield

$$\begin{aligned} \mathrm{H}_0 &= 2J + J' + \frac{1}{R} \\ \mathrm{H}_1 &= 2KS + K' + \frac{S^2}{R} \end{aligned} \tag{8.18}$$

where the following definitions hold:

$$\begin{aligned}
J &= -\frac{1}{R} + \exp(-2R)\left(1 + \frac{1}{R}\right) \\
J' &= \frac{1}{R} - \exp(-2R)\left(\frac{1}{R} + \frac{11}{8} + \frac{3}{4}R + \frac{1}{6}R^2\right) \\
K &= -\exp(-R)(1 + R) \\
K' &= -\exp(-2R)\left(-\frac{25}{8} + \frac{23}{4}R + 3R^2 + \frac{1}{3}R^3\right) + \frac{6}{R}(S^2(\gamma + \ln R) \\
&\quad + S'^2 E_i(-4R) - 2SS'E_i(-2R))
\end{aligned} \tag{8.19}$$

and

$$\begin{aligned}
S' &= \exp(R)\left(1 - R + \frac{1}{3}R^2\right) \\
\gamma &= \int_0^1 \frac{1 - \exp(-t)}{t}\mathrm{d}t - \int_1^\infty \frac{\exp(-t)}{t}\mathrm{d}t = 0.577\,215\,665\ldots \\
E_i(x) &= -\int_{-x}^\infty \frac{\exp(-t)}{t}\mathrm{d}t
\end{aligned} \tag{8.20}$$

where γ is the Euler constant and E_i is the exponential integral function. It is defined by taking the principal value of the integral.

The Heitler–London choice of the wave function results in a bound molecule as illustrated in Figure 8.2. The Heitler–London binding energy of 3.16 eV per molecule is too small when compared to a highly accurate calculation of 4.75 eV. The Heitler–London bond length is also too large, as one might have expected, given that their computed binding energy is too weak. The computed H—H bond length is 1.6 bohr units or about 0.85 Å compared to an accurate value of 0.74 Å [1].

Despite errors in the binding energy and bond length, Heitler and London's model was a notable success. They showed that quantum mechanics could account for the formation of a H$_2$ molecule. To do better, they needed better wave functions. As a simple example instead of assuming a wave function of the form $\psi = \exp(-r)$, they could have taken $\psi = \exp(-\alpha r)$ and then found $E(\alpha, R)$. At each value of R, an optimal value for α could be chosen. This improves the agreement notably. Using this approach, the binding energy increases to about 3.78 eV, a bit too low, but clearly better. The bond length is also improved. It decreases to 0.75 Å, essentially in exact agreement with the known value [2].

Earlier, we stated that the lowest energy state for the hydrogen molecule corresponded to a symmetric spatial function times an antisymmetric spin function. What about using an antisymmetric function for the spatial part and a symmetric spin function? Such a combination is also antisymmetric overall and obeys the Pauli principle. We use the following for the spatial part:

$$\psi(\vec{r}_2, \vec{r}_1) = \frac{[a(2)b(1) - a(1)b(2)]}{\sqrt{2(1 + S^2)}} \tag{8.21}$$

and a symmetric state for the spin. There are several ways of writing the spin state:

$$\begin{aligned}\chi(1,2) &= \chi_1(\uparrow)\chi_2(\uparrow)\\ \chi(1,2) &= [\chi_1(\uparrow)\chi_2(\downarrow) + \chi_1(\downarrow)\chi_2(\uparrow)]\\ \chi(1,2) &= \chi_1(\downarrow)\chi_2(\downarrow)\end{aligned} \tag{8.22}$$

Since the Hamiltonian we are considering has no spin, the energy of these three states is the same and corresponds to spin states with the total spin equal to −1, 0, or +1. In principle, there can be spin in the Hamiltonian, e.g. the electron spin can interact with a nuclear spin. We will not consider such situations.

It is easy figure out the energy of this triply degenerate state. (The nomenclature "degenerate" is odd. It means "equal" in physics and does not speak to the moral quality of the state, and it is never used as verb in this context.) All we need to do is change the "plus" sign in Eq. (8.16) to a minus sign:

$$E^*_{\mathrm{H}_2}(R) = -1 + \frac{\mathrm{H}_0 - \mathrm{H}_1}{1 - S^2} \tag{8.23}$$

where we label the high-energy state by $E^*_{\mathrm{H}_2}$. In Figure 8.2, we display the energies for the molecule using the spatial orbitals with symmetric and antisymmetric behavior. The "antisymmetric" energy, $E^*_{H_2}$, is less favorable than the "symmetric" energy, E_{H_2}, independent of the proton separation, R. An alternative nomenclature (which is widely used) is to refer to these states as *antibonding* and *bonding*, respectively.

8.2 General Basis

Even for the simplest molecule, an analytical solution for the Schrödinger equation is messy and only possible under special circumstances as with the Heitler–London choice of wave functions. Most approaches to solving the electronic structure problem are based on expressing the wave function as a sum of orbitals or basis functions. We outline the general strategy by considering a

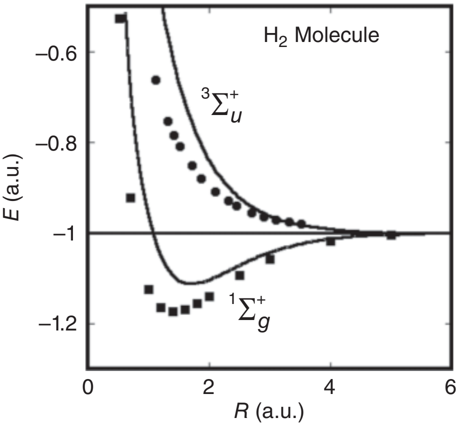


Figure 8.2 Energy of the H_2 molecule for a bonding state and an antibonding state. The spectroscopic notation is $^1\Sigma_g^+$ and $^3\Sigma_u^+$, respectively.

wave function within a one-electron picture. Our wave function is assumed to be composed of n basis functions:

$$\psi(\vec{r}) = \sum_{i=1}^{n} \alpha_i \, \phi_i(\vec{r}) \tag{8.24}$$

We insert this expression into the Schrödinger, or the Kohn–Sham, equation where for the moment we do not specify the details of the Hamiltonian:

$$H\psi(\vec{r}) = \sum_{i=1}^{n} \alpha_i \, H \, \phi_i(\vec{r}) = E\psi(\vec{r}) = E \sum_{i=1}^{n} \alpha_i \, \phi_i(\vec{r}) \tag{8.25}$$

Suppose the basis functions form an orthonormal set of functions. If they do not, we can usually make them so.

$$\int \phi_i^*(\vec{r})\phi_j(\vec{r}) \, \mathrm{d}^3 r = \delta_{ij} \tag{8.26}$$

Now we multiply both sides of Eq. (8.25) by ϕ_j^* and integrate:

$$\sum_{i=1}^{n} \alpha_i \int \phi_j^*(\vec{r}) H \, \phi_i(\vec{r}) \, \mathrm{d}^3 r = E \sum_{i=1}^{n} \alpha_i \int \phi_j^*(\vec{r})\phi_i(\vec{r}) \, \mathrm{d}^3 r = E\alpha_j \tag{8.27}$$

We can make this equation simpler by using the following notation:

$$H_{j,i} = \int \phi_j^*(\vec{r}) H \, \phi_j(\vec{r}) \, \mathrm{d}^3 r \tag{8.28}$$

We note $H_{ij} = Hij$, which is true if the matrix is real. If not, the matrix can be made so, at least for molecules. Using this notation, we can write

$$\sum_{i=1}^{n} (H_{ij} - E\delta_{i,j})\alpha_i \ = 0 \tag{8.29}$$

We have n such equations as we can take $j = 1, 2, \ldots, n$. For the sake of illustration, we take $n = 2$, which leaves us with two equations for the energy levels, E:

$$\begin{aligned} &(H_{11} - E)\alpha_1 + H_{21}\alpha_2 = 0 \qquad j = 1 \\ &H_{12}\alpha_1 + (H_{22} - E)\alpha_2 = 0 \qquad j = 2 \end{aligned} \tag{8.30}$$

This set of equations will only have a solution if the determinant of the coefficients vanishes; this is the condition for a "nontrivial" solution. A trivial solution would be $\alpha_1 = \alpha_2 = 0$, but this is not helpful as our wave function would be identical to zero! A "zero" wave function clearly cannot be normalized, so we discard that possibility forthwith.

We solve this equation for a nontrivial solution:

$$\det \begin{vmatrix} H_{11} - E & H_{12} \\ H_{21} & H_{22} - E \end{vmatrix} = 0 \tag{8.31}$$

We consider the simplest case first; we take $H_{11} = H_{22} = H_0$ and $H_{12} = H_{21} = H_1$. This yields

$$E_{\pm} = H_0 \pm H_1 \tag{8.32}$$

The respective wave functions correspond to

$$\psi_{\pm} = \frac{1}{\sqrt{2}}(\phi_1 \pm \phi_2) \tag{8.33}$$

This may seem like a very simplified example, but there is some "chemistry" here. Two orbitals interact and we obtain as many eigenvalues as we have orbitals. (This will always be true, i.e. we do not lose states. So the number of eigenvalues will be equal to the number of states in the basis.) One of the solutions involves a sum of the two orbitals, the other a difference between the two orbitals. This resembles a solution for the hydrogen molecule where we had "bonding" and "antibonding" states.

Suppose we are interested in a diatomic molecule with different atomic constituents (A,B) such as an AB diatomic molecule. We can easily accommodate this situation. We consider and use the following notation: $H_A = H_{11}$, $H_B = H_{22}$, and $H_{AB} = H_1$. This notation is clear in that if the atoms do not interact, then $H_{AB} = 0$ and we obtain the energy solution of $E = H_A$, $E = H_B$, the energy levels of the isolated atoms. We solve this equation for a nontrivial solution when the atoms interact:

$$\det\begin{vmatrix} H_A - E & H_{AB} \\ H_{AB} & H_B - E \end{vmatrix} = 0 \tag{8.34}$$

The solution is obtained by solving the binomial equation in E:

$$E = \overline{E} \pm \Delta E_{AB} \tag{8.35}$$

where $\overline{E}$ is the average of the isolated energy levels of the isolated atoms:

$$\overline{E} = \frac{1}{2}(H_A + H_B) \tag{8.36}$$

The interaction of the atoms leads to an energy difference between the lowest and the highest energy levels of

$$E_{\text{diff}} = 2\Delta E_{AB} = 2 \times \sqrt{\frac{(H_A - H_B)^2}{4} + H_{AB}^2} \tag{8.37}$$

Even though it is a simple model, we can use this formula to estimate the "ionicity" of the chemical bond. Suppose we take an energy scale such that our zero of energy is $\overline{E}$, in which case, $H_A = -H_B$. If we have a homopolar molecule, then $H_A = H_B$, which means in this energy scale $H_A = H_B = 0$. In contrast, if the molecule is heteropolar, then we can write $\Delta H = H_A - H_B$, where ΔH is a measure of the difference between the energy levels of the atoms A and B and can be used to define an ionic contribution. Using this notation,

$$E_{\text{diff}} = 2\Delta E_{AB} = 2 \times \sqrt{\frac{(\Delta H)^2}{4} + H_{AB}^2} \tag{8.38}$$

In the limit where $\Delta H \gg H_{AB}$, $E_{\text{diff}} \approx \Delta H$, i.e. the difference between the isolated energy levels. This limit corresponds to a molecule with very different constituents. In contrast, if $\Delta H \ll H_{AB}$, then we have the case where the constituent atoms are very similar, at least in terms of the isolated energy levels and $E_{\text{diff}} \approx 2|H_{AB}|$.

In the case of solids made of two constituent atoms, AB, where the sum of the valance electrons is 8, e.g. GaAs, ZnSe, and NaCl, this general description has been used to construct an ionicity scale by considering contributions to the energy gap between filled and empty states arising from homopolar (covalent) and heteropolar (ionic) contributions [3].

For a more realistic description, we need to consider the nature of the basis functions in Eq. (8.24). Standard descriptions of basis or orbital functions center on analytical functions such as Gaussians, exponentials, and plane waves. Gaussian and plane waves are popular choices as the matrix elements can often be analytically expressed. One might argue that Gaussian orbitals are more appropriate for molecules or clusters as they are localized whereas plane waves are infinite in extent and more sensible for crystals. Of course, one can take sums and differences of plane waves to produce a "localized" wave function. This requires a large number of plane waves, which can result in a computationally difficult problem. This complication is offset by the simplicity of plane waves and the corresponding ease in doing integrals.

8.2.1 Plane Wave Basis

A plane wave basis is a popular choice for crystalline systems. In crystalline matter, the atoms are periodically arranged. Consider a one-dimensional array of atoms, each atom the same and separated by a distance a. In this case, the potential also has a periodic behavior: $V(x + a) = V(x)$. The charge density of this system will reflect this symmetry. If $\rho(x)$ is the charge density of this system, then $\rho(x + a) = \rho(x)$. This symmetry enormously simplifies the problem as knowing the density in a part of the array allows us to know the charge density everywhere. The relevant part of the array is called a *unit cell*. In this very elementary example, the unit cell consists of the space defined within a distance of $\pm a/2$ about an atom.

Quantum mechanics is strange as we know. Quantum solutions associate a phase with the wave function; this is not a classical concept. Specifically, one might assume that a periodic wave function, i.e. $\psi(x + a) = \psi(x)$, also yields a charge density: $\rho(x + a) = |\psi(x + a)|^2 = |\psi(x)|^2 = \rho(x)$ with the correct symmetry. This form seems a satisfactory choice for a wave function.

However, consider a different choice, $\exp(i\lambda)\ \psi(x)$, where λ is a real number. This wave function yields exactly the same density, $\rho(x) = |\exp(i\lambda)\ \psi(x)|^2 = |\psi(x)|^2$, as $|\exp(i\lambda)|^2 = 1$ as multiplying a wave function by a phase factor such as $\exp(i\lambda)$ does not change the charge density. As we have noted before, in quantum mechanics some properties such as the phase of a wave function are deemed not observable, i.e. they cannot be quantified or measured. "Relative phases" can be measured and are important in some quantum phenomena, but this topic is beyond the scope of this discussion. The charge density is a different matter – unlike a wave function, charge density can be measured and observed.

There is a theorem by Felix Bloch that can handle situations where the potential is periodic. According to Bloch's theorem, in a one-dimensional periodic system the wave function can be assumed to change phase when translated by a distance a:

$$\psi(x + a) = \exp(ika)\psi(x) \tag{8.39}$$

The phase factor is now specified by a quantity ka. It is also possible to write Bloch's theorem in a different form:

$$\psi(x) = \exp(ikx)\, u(x) \tag{8.40}$$

where there is no phase factor associated with a periodic function: $u(x + a) = u(x)$. This allows us to write

$$\begin{aligned} \psi(x + a) &= \exp(ik(x + a))u(x + a) = \exp(ika) \exp(ikx)u(x + a) \\ &= \exp(ika)\, [\exp(ikx)\, u(x)] = \exp(ika)\psi(x) \end{aligned} \tag{8.41}$$

While one-dimensional periodic systems are useful as a teaching device, three-dimensional systems are more in tune with real physical systems. It is straightforward to generalize Bloch's theorem in one dimension to three dimensions. We take a potential that is periodic such that $V(\vec{r} + \vec{R}) = V(\vec{r})$ where $\{\vec{R}\}$ are *lattice vectors*. A lattice vector connects two equivalent points in a crystal. Lattice vectors can be defined as

$$\vec{R} = n_1\vec{a} + n_2\vec{b} + n_3\vec{c} \tag{8.42}$$

where $(\vec{a}, \vec{b}, \vec{c})$ define a unit cell, i.e. the parallelepipeds formed by the basis vectors, (n_1, n_2, n_3), are integers. For example, a cubic lattice (or periodic array of points) would have

$$\vec{a} = a\hat{x} \quad \vec{b} = a\hat{y} \quad \vec{c} = a\hat{z} \tag{8.43}$$

where $(\hat{x}, \hat{y}, \hat{z})$ are unit vectors and a is the spacing between lattice points. Bloch's theorem in three dimensions takes the form:

$$\psi_{\vec{k}}(\vec{r} + \vec{R}) = \exp(i\vec{k} \cdot \vec{R})\, \psi_{\vec{k}}(\vec{r}) \tag{8.44}$$

The *wave vector*, $\vec{k}$, labels each state. This is analogous to the case of a hydrogen atom. The spatial wave function of the atom is labeled by quantum numbers: (n, l, m). Here, the states are labeled by (k_x, k_y, k_z). The wave vector is related to momentum in the limit of a free electron.

Suppose we want to expand a wave function for a crystal in plane waves within Bloch's theory. Entire textbooks are written on this topic [4]. Here, we will sketch out the main issues.

For the moment, we consider a set of plane wave vectors with an unspecified set of plane waves: $\{\vec{G}\}$:

$$\psi_{\vec{k}}(\vec{r}) = \exp(i\vec{k} \cdot \vec{r}) \sum_{\vec{G}} \alpha(\vec{k}, \vec{G}) \exp(i\vec{G} \cdot \vec{r}) \tag{8.45}$$

where the sum is over this set of vectors, $\{\vec{G}\}$, and $\alpha(\vec{k}, \vec{G})$ are coefficients for the plane wave basis. This will work if Bloch's theorem holds:

$$\psi_{\vec{k}}(\vec{r} + \vec{R}) = \exp(i\vec{k} \cdot \vec{R})\, \psi_{\vec{k}}(\vec{r}) \tag{8.46}$$

Or, better if the following holds:

$$\sum_{\vec{G}} \alpha(\vec{k}, \vec{G}) \exp\left(i\vec{G} \cdot (\vec{r} + \vec{R})\right) = \sum_{\vec{G}} \alpha(\vec{k}, \vec{G}) \exp(i\vec{G} \cdot \vec{r}) \tag{8.47}$$

This clearly will hold for

$$\exp(i\vec{G}\cdot\vec{R}) = 1 \tag{8.48}$$

This condition defines the set of vectors $\{\vec{G}\}$, which are known as *reciprocal lattice vectors*. One can define a set of reciprocal vectors given by

$$\vec{G} = m_1\vec{A} + m_2\vec{B} + m_3\vec{C} \tag{8.49}$$

where the set $(\vec{A}, \vec{B}, \vec{C})$ generates a lattice in reciprocal space. These vectors are defined by

$$\vec{A} = \frac{2\pi}{\Omega}\vec{b}\times\vec{c} \quad \vec{B} = \frac{2\pi}{\Omega}\vec{c}\times\vec{b} \qquad \vec{C} = \frac{2\pi}{\Omega}\vec{a}\times\vec{b} \tag{8.50}$$

where Ω is the unit cell volume. Note that $\Omega = |\vec{a}\cdot\vec{b}\times\vec{c}|$ from elementary vector analysis. It is easy to show that

$$\begin{aligned}
&\vec{A}\cdot\vec{a} = 2\pi \quad \vec{A}\cdot\vec{b} = 0 \quad \vec{A}\cdot\vec{c} = 0\\
&\vec{B}\cdot\vec{a} = 0 \quad \vec{B}\cdot\vec{b} = 2\pi \quad \vec{B}\cdot\vec{c} = 0\\
&\vec{C}\cdot\vec{a} = 0 \quad \vec{C}\cdot\vec{b} = 0 \quad \vec{C}\cdot\vec{c} = 2\pi
\end{aligned} \tag{8.51}$$

By construction, these vectors have the property:

$$\vec{G}\cdot\vec{R} = 2\pi(n_1m_1 + n_2m_2 + n_3m_3) \tag{8.52}$$

and

$$\exp(\vec{G}\cdot\vec{R}) = \exp(2\pi(n_1m_1 + n_2m_2 + n_3m_3)) = 1 \tag{8.53}$$

as we desired. For a cubic lattice in real space, the reciprocal lattice vectors are

$$\vec{A} = \frac{2\pi}{a}\hat{x} \quad \vec{B} = \frac{2\pi}{a}\hat{y} \quad \vec{C} = \frac{2\pi}{a}\hat{z} \tag{8.54}$$

We can use reciprocal lattice vectors to directly express *any* periodic function. For example, we can write the potential as follows:

$$V(\vec{r}) = \sum_{\vec{G}} V(\vec{G})\exp(i\vec{r}\cdot\vec{G}) \tag{8.55}$$

where the Fourier coefficients, $V(\vec{G})$, can be determined by

$$V(\vec{G}) = \frac{1}{\Omega}\int V(\vec{r})\exp(-i\vec{r}\cdot\vec{G})\mathrm{d}^3r \tag{8.56}$$

This can be considered a Fourier series in three dimensions. The periodicity is evident by recognizing

$$\begin{aligned}
V(\vec{r}+\vec{R}) &= \sum_{\vec{G}} V(\vec{G})\exp(i(\vec{r}+\vec{R})\cdot\vec{G})\\
V(\vec{r}+\vec{R}) &= \sum_{\vec{G}} V(\vec{G})\exp(i\vec{r}\cdot\vec{G})\exp(i\vec{R}\cdot\vec{G}) = V(\vec{r})
\end{aligned} \tag{8.57}$$

Suppose we place some atoms in our unit cell. The unit cell is specified by the lattice vectors $\{\vec{R}\}$ and the atoms within the unit cell by a *basis*. In general, crystal structures are defined by (i) a unit cell and (ii) a basis. One considers a periodic array of lattice points and associates some atomic configuration with each lattice point. We can assign the atomic positions by considering a set of atoms specified by a set of vectors: $\{\vec{\tau}\}$.

Let us consider an elemental crystal where at each atomic site we place a potential, $V_a(\vec{r})$. The crystalline potential for such a crystal is given by

$$V(\vec{r}) = \sum_{\vec{R},\vec{\tau}} V_a(\vec{r} - \vec{R} - \vec{\tau}) \tag{8.58}$$

Using the definition given in Eq. (8.56), it is really easy to show that we can write this as

$$V(\vec{r}) = \sum_{\vec{G}} S(\vec{G}) V_a(G) \exp(i\vec{G} \cdot \vec{r}) \tag{8.59}$$

where $S(\vec{G})$ is called the *structure factor* and $V_a(G)$ is called the *form factor*. The structure factor is given by

$$S(\vec{G}) = \frac{1}{N_a} \sum_{\vec{\tau}} \exp(i\vec{G} \cdot \vec{\tau}) \tag{8.60}$$

where N_a is the number of atoms in the cell. The form factors are given by

$$V(G) = \frac{1}{\Omega_a} \int V_a(\vec{r}) \exp(i\vec{r} \cdot \vec{G}) d^3r \tag{8.61}$$

where Ω_a is the atomic volume, i.e. $\Omega = N_a\Omega_a$. We have made an additional assumption here. The atomic potential is taken to be spherically symmetric: $V_a(\vec{r}) = V_a(r)$. For many systems, this assumption is not a bad approximation. As a consequence, we can write $V_a(\vec{G}) = V_a(G)$.

Example 8.1 ***The Empirical Pseudopotential Method Applied to Diamond Structure Semiconductors*** We can illustrate a simplified approach using pseudopotentials. This approach is called the "empirical pseudopotential method" or EPM. The EPM provided the first realistic description of energy bands for semiconducting crystals and stands as a remarkable advance in understanding the electronic and optical properties of semiconductors [5, 6]. The method works very well for several reasons. First, the crystals in question are highly symmetric. This attribute reduces the number of degrees of freedom. In some cases, symmetry considerations result in a structure factor that vanishes, so the form factor is irrelevant since it is multiplied by zero. Second, pseudopotentials often quickly converge in reciprocal space. This allows us to terminate the sum over reciprocal vectors with a manageable number of terms. For elemental crystals such as silicon, the number can be just three.

Elements such as carbon, silicon, and germanium form in the diamond structure. The diamond structure is illustrated in Figure 8.3. Each atom is fourfold coordinated, i.e. each atom has four neighbors. The low coordination of such a structure is a "signature of quantum mechanics" as pair-wise forces cannot replicate such a structure, at least not physically "sensible" pairwise potentials. The forces holding the diamond crystal together are covalent bonds. One covalent

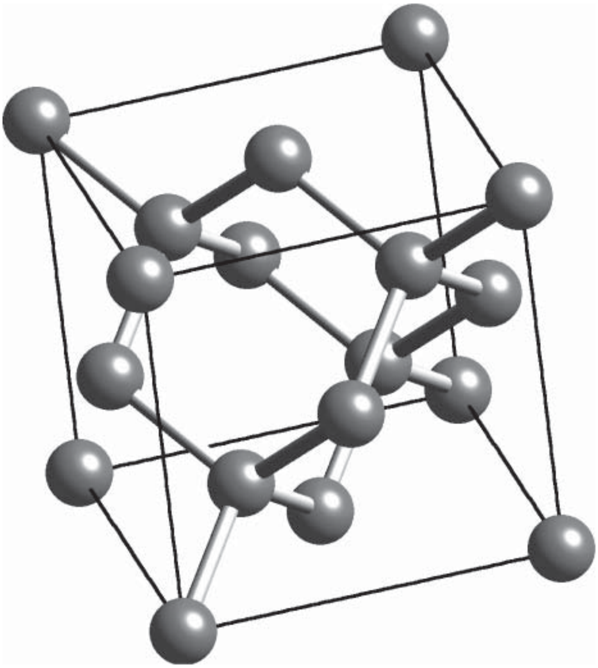


Figure 8.3 Structure of a diamond crystal.

bond holds two electrons and there are two bonds per atom (or four bonds per atom – with each bond shared by two atoms).

The structure of the diamond crystal can be constructed using unit cells that are either "conventional" or "primitive." The *conventional unit cell* of the diamond structure is cubic with the length of the edge given by the *lattice parameter*, *a*. Eight atoms are contained within the cell. The reciprocal lattice vectors are easy to represent because they form an orthogonal set. The "primitive cell" contains only two atoms, and would seem like a better choice, but the cell is not cubic. Nonetheless, it is the one commonly used in calculations as the number of states per cell is reduced by a factor of 4 and the computation load is notably reduced.

For a primitive unit cell, the lattice vectors are the same as for a face-centered cubic crystal.

$$\vec{a} = a(\hat{y} + \hat{z})/2 \qquad \vec{b} = a(\hat{x} + \hat{z})/2 \qquad \vec{c} = a(\hat{x} + \hat{y})/2 \tag{8.62}$$

If we look at the lattice points so generated and place an atom on each point, we would replicate the face-centered cubic crystal. The diamond crystal possesses the same lattice points, but associates two atoms with each lattice point. The basis for the diamond crystal can be chosen as

$$\tau_1 = -a(1,1,1)/8 \qquad \tau_2 = a(1,1,1)/8 \tag{8.63}$$

The combination of a lattice and a basis define a crystal structure.

We can generate the reciprocal lattice vectors for diamond using the formulae in Eq. (8.50). This yields the following:

$$\vec{A} = \frac{2\pi}{a}(-\hat{x}+\hat{y}+\hat{z}) \quad \vec{B} = \frac{2\pi}{a}(\hat{x}-\hat{y}+\hat{z}) \qquad \vec{C} = \frac{2\pi}{a}(\hat{x}+\hat{y}-\hat{z}) \tag{8.64}$$

This nonorthogonal triad of vectors complicates the description of reciprocal space. If we form reciprocal lattice vectors from

$$\vec{G} = n_1\vec{A} + n_2\vec{B} + n_3\vec{C} \tag{8.65}$$

we write the reciprocal lattice vectors, $\vec{G}$, for the diamond structure in Cartesian coordinates as

$$\vec{G} = \frac{2\pi}{a}(n, l, m) \tag{8.66}$$

where unlike a simple cubic crystal the indices (n, l, m) are not arbitrary integers. There are rules to be obeyed! It is easy to show that the integers must be either all odd or all even, e.g. $\vec{G} = \frac{2\pi}{a}(1, 1, 0)$ is not allowed, but $\vec{G} = \frac{2\pi}{a}(2, 0, 0)$ is permitted.

It is convenient to organize the $\vec{G}$-vectors for the diamond crystal by their magnitude squared in units of $(2\pi/a)^2$. In this scheme,

$$G^2 = (2\pi/a)^2[0, 3, 4, 8, 11, 12, \ldots] \tag{8.67}$$

are allowed magnitudes for the smallest vectors. Sometimes workers in the field just write $G^2 = 0, 3, 4, 8, \ldots$ understanding that there is an implicit prefactor of $(2\pi/a)^2$. To be specific, the smallest allowed value for G^2, which is not zero, comes from $\vec{G} = \frac{2\pi}{a}(1, 1, 1)$ and all sign permutations as the system is cubic.

The $\vec{G}$-space sets up the symmetry of the crystal in reciprocal space accompanied by the form factor, $S(\vec{G})$. The structure factor for the diamond structure is

$$S(\vec{G}) = \cos\left(\vec{G}\cdot\vec{\tau}\right) = \cos\left(\frac{\pi}{4}(l+m+n)\right) \tag{8.68}$$

For some values of G, this structure factor vanishes, e.g. if $\vec{G} = (2\pi/a)(2, 0, 0)$, then $\vec{G}\cdot\vec{\tau} = \pi/2$ and $S(\vec{G}) = 0$. If the structure factor vanishes, the corresponding form factor, $V(G)$, does not matter. Physically, contributions of these plane waves cancel out and cannot contribute to electronic potentials. In the case of diamond structure, this eliminates the $G^2 = 4, 12$ form factors. Also, the $G^2 = 0$ factor is not important for spectroscopy as $V(0)$ corresponds to the average potential and serves only to shift all the energy bands by a constant energy. For cohesive energies or ionization energies where an absolute reference is required, the shift is important.

The rapid convergence of the pseudopotential in Fourier space coupled with the vanishing of the structure factor for certain $\vec{G}$s means that only *three* form factors are required to fix the energy bands for diamond semiconductors such as Si and Ge: $V_\text{p}^\text{a}(G^2 = 3)$, $V_\text{p}^\text{a}(G^2 = 8)$, and $V_\text{p}^\text{a}(G^2 = 11)$. These form factors are treated as *adjustable parameters* and can be fixed by comparisons to reflectivity measurements or photoemission experiments [5]. In Table 8.1, we list empirical pseudopotential form factors for silicon and germanium.

Table 8.1 Empirical potentials.

		Potentials (Ry)		
Elements	**Lattice constant (Å)**	***V*(3)**	***V*(8)**	***V*(11)**
Si	5.43	−0.21	0.04	0.08
Ge	5.66	−0.23	0.01	0.06

We can use these parameters to solve for the electronic structure of Si and Ge crystals by solving an eigenvalue problem. We write

$$\left(\frac{-\hbar^2}{2m} + V_p(\vec{r})\right) \psi_{n\vec{k}}(\vec{r}) \;=\; E_n(\vec{k})\,\psi_{n\vec{k}}(\vec{r}) \tag{8.69}$$

where V_p is the pseudopotential expressed as in Eq. (8.59). We use the same approach as for the general basis problem. We expand the wave functions in a basis and utilize the orthonormal character of the basis to set up a secular equation. We write

$$\psi_{\vec{k}}(\vec{r}) = \exp(i\vec{k}\cdot\vec{r}) \sum_{\vec{G}} \alpha(\vec{k},\vec{G}) \exp(i\vec{G}\cdot\vec{r}) \tag{8.70}$$

and choose an element of the basis: $\exp(-i\vec{G}'\cdot\vec{r})$ and make use of orthonormality:

$$\frac{1}{\Omega}\int \exp(i(\vec{G}-\vec{G}')\cdot\vec{r})\,\mathrm{d}^3r = \delta_{\vec{G},\vec{G}'} \tag{8.71}$$

where Ω is the crystal volume. We use a plane wave basis in Eq. (8.69), multiply both sides of the equation by $\exp(-i\vec{G}'\cdot\vec{r})$, and integrate. The result is a secular equation of the form:

$$\det\left|\left(\frac{\hbar^2}{2m}\left(\vec{k}+\vec{G}\right)^2 - E\right)\delta_{\vec{G},\vec{G}'} + S(\vec{G}-\vec{G}')V_\mathrm{a}\left(\vec{G}-\vec{G}'\right)\right| = 0 \tag{8.72}$$

There are standard methods for determining the eigenvalues or energy levels for this determinant. We will not discuss them here, but in the following section.

There is one subtle point. Namely, the set of $\{\vec{G}\}$ is terminated by the matrix size so that $|\vec{k}+\vec{G}|^2$ is less than some specified cutoff: Q_max. This cutoff defines the overall accuracy of the solution. For a material such as Si or Ge, the pseudopotential form factors decay rapidly and one can use $Q_\mathrm{max} \approx (2\pi/a)\times 4$ and calculate an accurate energy band structure.

The solution of Eq. (8.72) yields the eigenvalues for any given $\vec{k}$-point. These eigenvalues are commonly labeled by an index n where $n = 1, 2, 3, \ldots$ and the $\vec{k}$, i.e. the eigenvalues are labeled $E_n(\vec{k})$. Since these eigenvalues as a function of $\vec{k}$ form continuous levels, they are called *energy bands,* where n is called the band index. Owing to the three-dimensional character of $E_n(\vec{k})$, displaying the bands becomes problematic. Typically one chooses high symmetry directions and plots the corresponding bands. The energy band structure

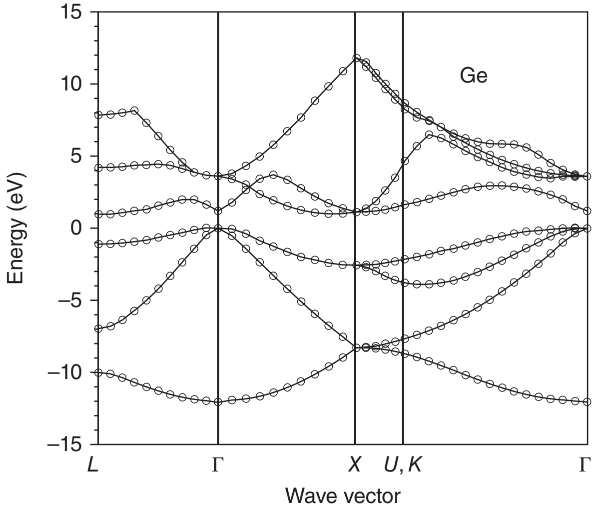


Figure 8.4 Energy band structure for the Ge crystal. The energy zero is taken to be the valence band maximum. Only the bands near the band gap are shown.

for germanium is shown in Figure 8.4. The discrete nature of the calculated eigenvalues is shown by the dots. If we take more wave vectors along the symmetry directions, these dots would eventually merge to form an energy band.

In Figure 8.4, several directions for $\vec{k}$ are shown along high symmetry directions. For historical reasons, Greek and Roman letters are used to mark high symmetry points for $\vec{k}$. The point $\vec{k} = 0$ is called "Γ," the $\vec{k} = (2\pi/a)(1/2, 1/2, 1/2)$ point is called "L," the $\vec{k} = (2\pi/a)(1, 1, 1)$ point is called "X," and the $\vec{k} = (2\pi/a)(3/4, 3/4, 0)$ point is called K. Thus, points from Γ to L are along the (111) direction and points from Γ to X are along the (100) direction.

The electronic structure of a solid such as germanium is very different than a germanium atom in which discrete levels occur. In crystalline germanium, there are energy levels for a given $\vec{k}$. These levels will vary with the wave vector as shown in the figure. Each eigenvalue solution can hold two electrons. There are a total of eight electrons in the unit cell and four occupied eigenstates or energy bands. In Ge, there are four filled energy bands, an energy gap, and then empty bands. The filled or occupied bands are called *valence bands* and the empty states are called *conduction bands.*

Insulators or semiconductors have an energy gap between the valence and conduction bands called the "band gap." The size of the band gap determines whether a crystal is an insulator or a semiconductor. If the band gap is small, typically less than a few electronvolts, the crystal is a semiconductor; otherwise, the crystal is an insulator. For example, germanium has a band gap of approximately 0.7 eV, so it is a semiconductor. Since most of the interesting optical and transport properties occur near the top of the valence band and bottom of the conduction band, often the energy bands are shown in more detail near the gap. This is illustrated in Figure 8.5.

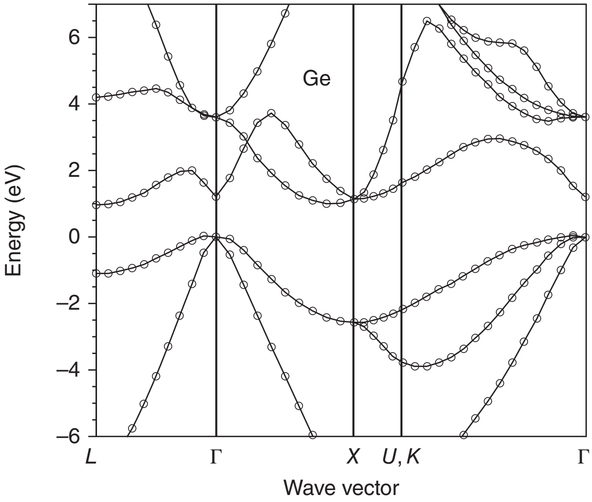


Figure 8.5 Energy band structure for the Ge crystal. The energy zero is taken to be the valence band maximum. Only the bands near the band gap are shown.

An insulator such as quartz (silicon dioxide) has a band gap of over 8 eV. Owing to the large band gap quartz is transparent. Visible light is composed of photons with roughly 1.5–3 eV. The absorption of a photon from visible light is not sufficient to excite an electron from an occupied state to an empty state and the photon passes unabated through the quartz crystal. In contrast, a similar photon will be absorbed by a germanium crystal. Germanium crystals are opaque and appear as metallic gray.

Accurate calculations of energy band structure for crystals are very useful guides for understanding optical and dielectric properties, especially for electronic materials. The success of the EPM is remarkable and stands as one of the notable achievements in the study of condensed matter. Without the EPM, our understanding of semiconducting materials would be likely to be set back by a generation. However, our focus here is not on the energy bands of solids per se. There are superb references in the literature for energy bands, especially for semiconductors [6].

8.2.2 Plane Waves Applied to Localized Systems

Plane wave based methods for energy bands of electronic materials such as silicon and germanium are easy to implement and widely applicable. Coupled to the EPM, they lead to the realistic energy band structures for semiconductors and simple metals, and can even be applied to transition metals. However, there are issues with plane waves for systems that lack periodicity. The simplicity of Bloch's theorem no longer holds, and $\vec{k}$ is no longer a good quantum number. Moreover, a basis of plane waves is infinite in extent and appears to be a poor choice to describe a finite or localized system such as a molecule or a cluster of atoms. So one might ask, why did we choose to emphasize this approach?

Fortunately, there is a workaround that allows one to continue using plane waves to form a basis [7]. Suppose we are interested in a small cluster of silicon, i.e.

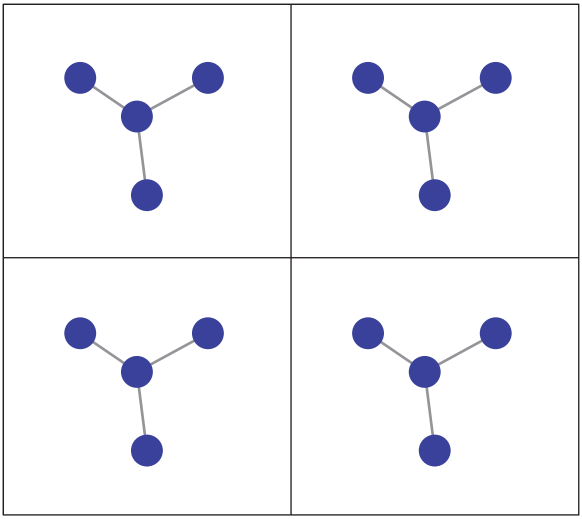


Figure 8.6 Supercell geometry for an atomic cluster.

several atoms of silicon acting in isolation. We place this cluster in a large cell (a "supercell") and surround it by vacuum. We then periodically repeat the cell and fill up all space. We now have a *periodic* array of clusters. Even though the periodicity is an artifact, we can now use the same tools as for a crystal. A schematic of this geometric construction is presented in Figure 8.6. If the vacuum region is sufficiently large, the clusters will not interact with each other. A solution to this set of problems corresponds to a periodic array of *isolated* clusters and the electronic structure of this system will correspond to that of an isolated cluster. Bloch's theorem will hold, although the wave vector $\vec{k}$ will no longer have any meaning.

There are some notable advantages to this method. All the tools developed for solving for the electronic structure of a crystal are at our disposal and the method is easy to implement. We just have to be careful to leave enough empty space around the system of interest.

There are a few drawbacks in using this method. If the empty space or selvage region between the molecules or clusters is not sufficiently large, there will be an interaction between neighboring cells. This results in the energy levels for the molecule or cluster having dispersion, i.e., the energy levels become energy "bands." This can always be tested by examining the energy levels at different $\vec{k}$ points. If the dispersion is too large, the cells can be made larger. Of course, a larger cell requires more plane waves to converge. This is wasteful as the extra plane waves are only used to describe the free space. There is another subtle reason for not wanting to use supercells. Suppose we want to charge the isolate system; e.g. suppose we want to compute the electron affinity of the molecule by adding an extra electron. If the supercell is not a neutral object, the electrostatic interactions of the cell with other cells will diverge. There are prescriptions for removing the divergences, but it is better to avoid them from the start. A technical drawback also exists. When we solve the electronic structure with plane waves, various matrix elements are required. These elements often represent the Fourier transform of a wave function or potential term. Fast Fourier transforms can be used to evaluate the wave function or potential. One might think that such transforms are "fast" as the name implies. This need not be true. Modern computers are slowed down by "global" communications that occur with plane waves.

8.3 Solving the Eigenvalue Problem

In our previous examples, we solved an "eigenvalue" problem. These problems can be written as

$$\mathscr{H}\,\psi_n = E_n\psi_n \tag{8.73}$$

There are several ways to solve eigenvalue problems. We consider the following procedure. Suppose we could expand the eigenvalues as follows:

$$\psi_n = \sum_{j=1}^{N} a_{n,j}\phi_j \tag{8.74}$$

where ϕ_j is a "basis," e.g. a set of Gaussians or plane waves. Let us assume (yet again) the following:

$$\int \phi_i^*\phi_j \,\mathrm{d}^3r = \delta_{ij} \tag{8.75}$$

This is usually the case. Assuming that the functions are normalizable, we can always make them so. We can also make sure that they are orthogonal using standard methods. We used a similar procedure to illustrate the variational theorem. Using the wave functions in Eq. (8.73), we have the following:

$$\mathscr{H}\,\sum_{j=1}^{N} a_{n,j}\phi_j = E_n\,\sum_{j=1}^{N} a_{n,j}\phi_j \tag{8.76}$$

We now multiply both sides by ϕ_k^* and integrate. We get the following:

$$\sum_{j=1}^{N} a_{n,j}\mathscr{H}_{kj} = E_n\,a_{n,k} \tag{8.77}$$

where

$$\mathscr{H}_{kj} = \int\, \phi_k^*\mathscr{H}\,\phi_j\,\mathrm{d}^3r \tag{8.78}$$

and we have made use of the orthogonality of the wave function basis. Eq. (8.77) can be recast as a matrix equation:

$$H\Psi = \Psi\Lambda \tag{8.79}$$

The matrix elements of the Hamiltonian matrix, H, are such that $H_{ij} = H_{ji}^*$, i.e. the complex conjugate of the transposed matrix element is invariant. Such matrices are called "Hermitian" and this property ensures that the eigenvalues E_n are real (this is not hard to prove). In most cases, we will have matrix elements that are real, which implies that the matrix is symmetric. For the wave function matrix, Ψ, we have

$$\Psi = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{N1} \\ a_{12} & a_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N} & \cdots & \cdots & a_{NN} \end{pmatrix} \tag{8.80}$$

Each column represents the coefficients for a given eigenvalue where the matrix Λ represents a diagonal matrix containing the corresponding eigenvalues:

$$\Lambda = \begin{pmatrix} E_1 & 0 & \dots & 0 \\ 0 & E_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & E_N \end{pmatrix} \tag{8.81}$$

The size of the matrix, $N \times N$, is determined by the number of eigenvalues. For example, if we expand the wave function in 100 plane waves, the matrix would be 100×100.

There are several ways to solve such a matrix equation. Let us consider the case for a two by two matrix, i.e. one in which there are only two eigenvalues.

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} E_1 & 0 \\ 0 & E_2 \end{pmatrix} \tag{8.82}$$

This can be split into two equations:

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} = E_1 \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} \tag{8.83}$$

and

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = E_2 \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} \tag{8.84}$$

We have handled this before in Eq. (8.38). For a nontrivial solution we need to have the following determinant vanish:

$$\det \begin{vmatrix} H_{11} - E & H_{12} \\ H_{21} & H_{22} - E \end{vmatrix} = 0 \tag{8.85}$$

This will yield two roots corresponding to E_1 and E_2. Once we know the roots we can solve for the coefficients, a_{ij}. For example, if we know E_1, then

$$a_{11} = \frac{-H_{12}}{H_{11} - E_1} a_{12}$$

This condition, plus a normalization condition, $a_{11}^2 + a_{12}^2 = 1$, is sufficient to determine the expansion coefficients.

For a general case, the eigenvalue problem may involve matrices of notable size, e.g. if we want to know the eigenvalues for a 10 000 atom cluster of silicon, we might need to find 20 000 eigenvalues. Here we illustrate the essential features of a powerful method, called the "power method," to solve for the eigenvalues.

Suppose we knew the eigenfunctions; we label them as ψ_j and consider an approximate solution for the largest eigenvalue, E_1. We will assume bound states so that the magnitude of E_1 is the largest (negative) eigenvalue we seek. We expand the corresponding wave function ϕ_1 as follows:

$$\phi_1 = \sum_{j=1}^{N} c_j \psi_j \tag{8.86}$$

We operate on this state with the Hamiltonian matrix:

$$H\phi_1 = \sum_{j=1}^{N} c_j \, H \, \psi_j = \sum_{j=1}^{N} c_j \, E_j \, \psi_j \tag{8.87}$$

where we have used $H\psi_n = E_n \psi_n$. We factor out the largest eigenvalue:

$$H\phi_1 = E_1 \left(c_1 \psi_1 + \sum_{j=2}^{N} c_j \, (E_j/E_1) \psi_j \right) \tag{8.88}$$

Suppose we define a new eigenvector:

$$\phi_1^{(1)} = \left(c_1 \psi_1 + \sum_{j=2}^{N} c_j \, (E_j/E_1) \psi_j \right) \tag{8.89}$$

We can repeat this operation again using the new eigenvector:

$$H\phi_1^{(1)} = E_1 \left(c_1 \psi_1 + \sum_{j=2}^{N} c_j \, (E_j/E_1)^2 \, \psi_j \right) \tag{8.90}$$

Suppose we keep repeating this operation using an updated eigenvector and we do the operation k times:

$$H\phi_1^{(k)} = E_1 \left(c_1 \psi_1 + \sum_{j=2}^{N} c_j \, (E_j/E_1)^k \, \psi_j \right) \tag{8.91}$$

Since E_1 is the largest eigenvalue $E_j/E_1 < 1$ and $(E_j/E_1)^k \to 0$ for all $j \neq 1$ in the limit of large values of k. In this limit, we have

$$H\phi_1^{(k)} = E_1 \phi_1^{(k)} \tag{8.92}$$

In short, if we keep applying H to an approximate eigenvector, we can eventually extract the largest eigenvalue. This method is known as the "power method." It is a practical approach to finding eigenvalues.

This is all well and good, provided we only want the largest eigenvalue. However, in general we want more, a lot more. We can use a simple operation for doing this. Suppose we wish to compute the second largest eigenvalue, which we will label E_2. We can do the following. We orthogonalize our approximate wave function to the lowest one. Let us take

$$\phi_2 = \sum_{j=1}^{N} b_j \psi_j \tag{8.93}$$

However, we now orthogonalize this eigenfunction so that

$$\tilde{\phi}_2 = \phi_2 - (\psi_1^{\mathrm{T}} \phi_2) \psi_1 \tag{8.94}$$

We operate with H on this eigenstate to obtain

$$H\tilde{\phi}_2 = H\phi_2 - (\psi_1^{\mathrm{T}}\phi_2)H\psi_1 = \sum_{j=1}^{N} b_j E_j \psi_j - \left(\psi_1^{\mathrm{T}} \sum_{j=1}^{N} b_j \psi_j\right) E_1 \psi_1 \tag{8.95}$$

which can be arranged to yield

$$H\tilde{\phi}_2 = b_1 E_1 \psi_1 + \sum_{j=2}^{N} b_j E_j \psi_j - b_1 E_1 \psi_1 = E_2\left(b_2\psi_2 + \sum_{j=3}^{N} b_j (E_j/E_2)\psi_j\right) \tag{8.96}$$

We have made use of $(\psi_1^{\mathrm{T}}\psi_j) = 0$. We illustrated the orthogonality in Chapter 6; see Eq. (6.5). If the eigenstates are degenerate, they can also be made orthogonal to each other. As before, if we apply H to generate new eigenstates k times, then we get the following:

$$H\tilde{\phi}_2 = E_2\left(b_2\psi_2 + \sum_{j=3}^{N} b_j (E_j/E_2)^k \psi_j\right) \tag{8.97}$$

Since the sum starts with $j = 3$, we have $(E_j/E_2)^k \rightarrow 0$ for large k as before. In this case,

$$H\tilde{\phi}_2 = E_2\tilde{\phi}_2 \tag{8.98}$$

We can keep doing this operation until we have as many eigenstates as we want, i.e. we keep orthogonalizing the state of interest to the lower energy state and operate on the state until the desired eigenvalue is obtained.

8.3.1 An Example Using the Power Method

We can illustrate the operation here with an example. We take the following problem:

$$H = \begin{pmatrix} -6 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & -4 \end{pmatrix}$$

We choose a "random" initial state:

$$\phi_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

We then take the product of $H\phi_1$:

$$\phi_1^{(1)} = \begin{pmatrix} -5 \\ -2 \\ -3 \end{pmatrix}$$

We note that if ϕ_i is an eigenstate, then $c\phi_1$ is also one. It makes it easy to read off the eigenvalue by always normalizing the first element to unity. We do this to get

$$\phi_1^{(1)} = \begin{pmatrix} 1 \\ 2/5 \\ 3/5 \end{pmatrix}$$

and find $H\phi_1^{(1)}$:

$$\phi_1^{(2)} = \begin{pmatrix} -18/5 \\ 0 \\ -2 \end{pmatrix}$$

Again, we normalize the next eigenstate so that

$$\phi_1^{(2)} = \begin{pmatrix} 1 \\ 0 \\ 5/14 \end{pmatrix}$$

Operating by H yields

$$\phi_1^{(3)} = \begin{pmatrix} -6 \\ 19/14 \\ -10/7 \end{pmatrix}$$

To summarize, the first operation of H produced -5 and the second operation produced $-5\frac{3}{5}$ and the third yielded -6. We can keep this up, but the fractions start getting messy. After about ten applications, we get $E_1 \approx -6.47$. The exact value is $E_1 = -6.48$. The convergence is not very rapid in this case, but the operations are easy to compute. Once we know the eigenvalue, we can find the eigenstate. After a number of iterations, we would find

$$\phi_1 = \begin{pmatrix} 1 \\ -0.4811 \\ 0.1939 \end{pmatrix}$$

We can now use this state to find the next eigenvalue, E_2. We orthogonalize our initial guess to this state:

$$\tilde{\phi}_2 = \phi_2 - (\phi_1^{\mathrm{T}}\phi_2)\phi_1$$

When we do this operation, we need to make sure that $(\phi_1^{\mathrm{T}}\phi_1) = 1$. We renormalize ϕ_1:

$$\phi_1 = \begin{pmatrix} 0.8877 \\ -0.4271 \\ 0.1721 \end{pmatrix}$$

We can use the same initial guess for ϕ_2 as before. But we can choose a different state if we like. We take

$$\phi_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

We do this so that our guess is orthogonal to our original guess, but this is not necessary. Any guess will do as long as the guess does not correspond to ϕ_1! We find our $\tilde{\phi}_2$ from

$$\tilde{\phi}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} - \begin{pmatrix} 0.8877 & -0.4271 & 0.1721 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \begin{pmatrix} 0.8877 \\ -0.4271 \\ 0.1721 \end{pmatrix} = \begin{pmatrix} 0.5319 \\ 0.7441 \\ -0.8969 \end{pmatrix}$$

We renormalize this (yet again) so that the lead term is unity:

$$\tilde{\phi}_2 = \begin{pmatrix} 1 \\ 1.3989 \\ -1.6862 \end{pmatrix}$$

We then form $H\tilde{\phi}_2^{(1)}$ to get

$$H\tilde{\phi}_2^{(2)} = \begin{pmatrix} -4.6012 \\ -6.2818 \\ 8.1436 \end{pmatrix}$$

We then normalize this function to

$$\tilde{\phi}_2^{(3)} = \begin{pmatrix} 1 \\ 1.3653 \\ -1.7699 \end{pmatrix}$$

This update is still orthogonal to the known solution ϕ_1. Again we form the product, $H\tilde{\phi}_2^{(3)}$, to get

$$H\tilde{\phi}_2^{(3)} = \begin{pmatrix} -4.6347 \\ -6.2309 \\ 8.4448 \end{pmatrix}$$

After repeating this operation and checking the orthogonalization, the eigenvalue for E_2 will converge to −4.69. This is a slow operation. Our first estimates for E_2 were −4.60 and −4.63. The correct and normalized value for ϕ_2 is

$$\phi_2 = \begin{pmatrix} -0.3971 \\ -0.5207 \\ 0.7558 \end{pmatrix}$$

We perform one more operation to estimate E_3. For $\tilde{\phi}_3$, we want to make sure our initial guess is orthogonal to ϕ_1 and ϕ_2. This way, the power method will not converge to either of the first two eigenvalues. We again take an initial guess for ϕ_3 and perform the following operation:

$$\tilde{\phi}_3 = \phi_3 - (\phi_1^{\mathrm{T}}\phi_3)\phi_1 - (\phi_2^{\mathrm{T}}\phi_3)\phi_2$$

We take our initial guess to be

$$\phi_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

This yields our first attempt for $\phi_3^{(1)}$ (after normalizing the first element to unity as before):

$$\phi_3^{(1)} = \begin{pmatrix} 1 \\ 3.1708 \\ 2.7100 \end{pmatrix}$$

Forming the product $H\phi_3^{(1)}$ yields

$$H\phi_3^{(1)} = \begin{pmatrix} -2.8296 \\ -8.9730 \\ -7.6693 \end{pmatrix}$$

Our first estimate for E_3 is −2.83. The correct value is −2.93. The reader can guess why this might be so. We only have so many degrees of freedom and by insisting the eignvector ϕ_3 be orthogonal to ϕ_1 and ϕ_2 we use them up!

The power method lies at the heart of most modern electronic structure codes. The repeated application of H on an eigenstate is a standard operation, which is easily implemented on modern computational platforms.

Of course, the method is not as easy as we have implied. In our example, the largest eigenvalue was the most negative. Suppose we had a system where the largest eigenvalue is not the one we want; e.g. suppose we consider a system where a large positive eigenvalue exists. There are methods for finding such eigenvalues. Most use a method related to the power method illustrated here [8].

References

1 Heitler, W. and London, F. (1927). Wechselwirkung neutraler Atome und homopolare Bindung nach der Quantenmechanik. *Zeitschrift fur Physik* 44: 455.

2 Chen, G., Chin, S.A., Dou, Y. et al. (2005). The two-electron molecular bond revisited: from Bohr orbits to two-center orbitals. In: *Advances in Atomic, Molecular, and Optical Physics*, vol. 51 (ed. H.H. Stroke), 93. Elsevier.

3 Phillips, J.C. (1970). Ionicity of the chemical bond in crystals. *Review of Modern Physics* 42: 317.

4 Kittel, C. (2005). *Introduction to Solid State Physics*, 8e. Wiley.

5 Cohen, M.L. and Bergstresser, T.K. (1966). Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures. Physical Review 141: 789.

6 Cohen, M.L. and Chelikowsky, J.R. (1989). *Electronic Structure and Optical Properties of Semiconductors*, 2e. Springer-Verlag.

7 Cohen, M.L., Schlüter, M., Chelikowsky, J.R. and Louie, S.G. (1975). Self Consistent pseudopotential method for localized configurations: Molecules. *Physical Review B* 12: 5575.

8 Cheney, W. and Kincaid, D. (2004). *Numerical Mathematics and Computing*. Thomson Learning.

9

Engineering Quantum Mechanics

> *I'll bet any quantum mechanic in the service would give the rest of his life to fool around with this gadget.*
>
> — Chief Engineer Quinn in the *Forbidden Planet*

Understanding and designing structural materials is based on Newton's laws. These laws concern the study of statics – forces, constraints, equilibria, bridge building – and are standard fare for mechanical engineers. Likewise, designing electronic applications and devices is largely based on Maxwell's equations. These questions cover the study of electronics – conduction, transport, electronic motors, and switches – and are standard fare for electrical engineers. However, the engineering of materials based on quantum mechanics is different. Typically, universities do not have Departments of Quantum Mechanical Engineering as they have Departments of Mechanical Engineering or Departments of Electric Engineering. This situation is largely because scientists and engineers have yet to fully address Dirac's challenge so that they can routinely apply "practical methods of quantum mechanics" to design new materials and applications thereof.

In this chapter, we introduce a "practical" method for applying quantum mechanics to localized systems such as atoms, molecules, clusters, and nanoparticles. These practical methods are not at the level of those developed for, say, mechanical engineering. Yet, the pathway is at least partially illuminated for progress in this area. Our focus will be on methods that can be implemented using simple computational algorithms that can be executed using modest computational means, e.g. a laptop computer.

9.1 Computational Considerations

We have stressed that the combination of pseudopotentials and density functional theory offers a workable methodology for computing the electronic structure and energetics of materials.

We focus on engineering methods to solve the Kohn–Sham equation:

$$\left(\frac{-\hbar^2 \nabla^2}{2m} + V_{\mathrm{p}}^{\mathrm{ion}}(\vec{r}) + V_{\mathrm{H}}(\vec{r}) + V_{\mathrm{xc}}[n(\vec{r})] \right) \phi_i(\vec{r}) = E_i \phi_i(\vec{r}) \tag{9.1}$$

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

where the notation is the same as in Eq. (7.33), except that we have replaced the nuclear potential with the ionic pseudopotential. The Kohn–Sham equation is solved by a self-consistent cycle as illustrated in Figure 7.3. The only difference here is that the nuclear potential is now an ionic pseudopotential.

We have indicated in previous chapters how this equation might be solved for a crystal. We could expand the wave functions in a basis, e.g. plane waves, set up a secular equation, and solve the resulting eigenvalue problem. For crystalline matter, this makes good sense as we illustrated in Chapter 8 (see Figure 8.5) as the periodic nature of signs and cosines can be used to reflect the translational symmetry of a crystal.

For non-crystalline matter such as a cluster, the use of plane waves is more problematic. While plane waves are easy to implement, which is a great asset, the basis is infinite in extent whereas a cluster is finite. This is an inherent mismatch in length scales, which has some notable negative consequences such as requiring the use of many plane waves to describe the potential in a vacuum region.

Still, one can use plane waves for a cluster or some other finite system. We indicated a viable approach by placing the cluster of interest in a large supercell, which retains periodicity albeit in an artificial way. This approach, while simple, has several physical downsides as in having spurious interactions between neighboring cells (see Section 8.2.2). One notable downside is "computational." Consider a large cluster. Suppose we modify the position of an atom on the exterior of a nanostructure. A plane wave description of this modification would require us to consider modifications in the basis over the entire system as the basis is infinite in extent. Global communications can notably reduce computational efficiency in contemporary computers.

Localized bases are more efficient as they reduce communications. The basis around a particular atom can be modified without changing the basis on distant, noninteracting atoms. However, localized bases have their drawbacks. They are often difficult to implement. When an atom is moved, the localized basis moves with the atom. Forces due to changes in the convergence as the atoms are moved are called Pulay forces [1]. The basis needs to be rechecked for convergence every time the structure is altered, although in practice this is rarely done. While localized bases result in smaller matrices, the matrix elements can be complicated to compute and the resulting matrix, while small, is dense.

We focus on *real-space* methods for these reasons. Real-space methods do not require the need for global communications, similar to a plane wave basis. In real-space methods, the "basis" is not infinite in extent and does not require the use of supercells, *i.e.* charged clusters can be examined without the use of a compensating background.

However, the matrix is large in real-space methods. Suppose we consider a cube and take 100 divisions along each side. We have 1 000 000 subdivisions. If we characterize each division by a grid point in its center, we would have to solve a matrix that is 1 000 000 by 1 000 000. This sounds bad, but it is not! The matrix is extraordinarily sparse and is never fully stored – only the nonzero matrix elements are.

Real-space solutions of the Kohn–Sham equation can be accomplished using standard methods such as "finite difference" [2], or "finite element" solutions [3]. Finite element methods are often used for situations where the boundary conditions are complex; however, they tend to be more difficult to implement than high-order finite difference methods. "High-order" finite difference methods are well suited for our applications owing to their ease of implementation and simplicity. Our boundary conditions are not complex in most cases. The wave function in question either vanishes or is periodic at the boundary.

9.2 Finite Difference Methods

Suppose we have knowledge of a function on a grid. We consider a one-dimensional grid where we know the value a function, $f(x_j)$, on a set of grid points, $\{x_i\}$, which are uniformly spaced, $h = x_{j+1} - x_j$, for all values of j. If the function is well behaved in the usual sense – no poles and no discontinuities (a kind and gentle function wherein a Taylor series would work) – then we make a standard approximation for its second derivative:

$$\left(\frac{d^2 f}{dx^2}\right)_{x_i} \approx \frac{f(x_{i+1}) + f(x_{i-1}) - 2f(x_i)}{h^2}$$

The error we make in this expression scales as $\mathcal{O}(h^2)$. We can do better by including more points:

$$\left(\frac{d^2 f}{dx^2}\right)_{x_i} \approx \frac{-f(x_{i+2}) + 16f(x_{i+1}) - 30f(x_i) + 16f(x_{i-1}) - f(x_{i-2})}{12h^2}$$

This expression is accurate to $\mathcal{O}(h^4)$.

In general, we can write

$$\left(\frac{d^2 f}{dx^2}\right)_{x_i} \approx \frac{1}{h^2} \sum_{n=-N}^{N} C_n f(x_i + nh)$$

which is accurate to $\mathcal{O}(h^{2N})$. Coefficients up to $2N$ are given in Table 9.1.

One can increase the order and get better expressions, but the increase in accuracy comes at the cost of knowing the wave functions at more points. The accuracy of the finite order expression is ultimately limited by the number of digits carried by the computational platform of interest, e.g. anything above $N \approx 12$ is apt to suffer from round off errors.

Table 9.1 Expansion coefficients for higher order finite difference expressions [4].

Order $\mathcal{O}(h^{2N})$	C_0	$C_{\pm1}$	$C_{\pm2}$	$C_{\pm3}$	$C_{\pm4}$	$C_{\pm5}$	$C_{\pm6}$
$N = 1$	−2	1					
$N = 2$	−5/2	4/3	−1/12				
$N = 3$	−49/18	3/2	−3/20	1/90			
$N = 4$	−205/72	8/5	−1/5	8/315	−1/560		
$N = 5$	−5269/1800	5/3	−5/21	5/126	−5/1008	1/3150	

Higher order difference methods are particularly well suited for the problem at hand, particularly when using pseudopotential methods. The pseudopotential wave functions in the case of an atom are nodeless and "smoothly" varying. Unlike the all-electron potential, the pseudopotential has no singularity at the nuclear site and the wave function lacks a cusp at the nuclear site.

We can discretize the Kohn–Sham equation in three-dimensional Cartesian coordinates:

$$\begin{aligned}
-\frac{\hbar^2}{2m}&\left[\sum_{n_1=-M}^{M} C_{n_1}\psi_n(x_i+n_1h,y_j,z_k)+\sum_{n_2=-M}^{M} C_{n_2}\psi_n(x_i,y_j+n_2h,z_k)\right.\\
&\left.+\sum_{n_3=-M}^{M} C_{n_3}\psi_n(x_i,y_j,z_k+n_3h)\right]+\left[\sum_{lmn,NL} V_{\text{ion}}^{\text{p}}(x_l,y_m,z_n)\right.\\
&\left.+V_{\text{H}}(x_i,y_j,z_k)+V_{\text{xc}}(x_i,y_j,z_k)\right]\psi_n(x_i,y_j,z_k)=E_n\quad\psi_n(x_i,y_j,z_k),\qquad(9.2)
\end{aligned}$$

where the variable of interest is the wave function, $\psi_n(x_i, y_j, z_k)$, at a grid point (x_i, y_j, z_k). In this example, we have taken the grid points to be uniformly spaced. Uniformly spaced points are not mandatory, but to do otherwise enormously complicates the problem. We avoid them even when they may offer some reduction in computational time.

A typical geometry for a cluster is illustrated in Figure 9.1. The Kohn–Sham wave functions are defined within the domain and taken to be zero outside of the domain. This simple boundary condition can be checked by changing the domain size to ensure that the results are not altered.

An obvious point concerns an interplay between the grid spacing, h, and the order of the finite differencing, M. A finer grid with a low-order finite difference can give similar accuracy to a coarser grid with a higher order finite difference. Communications are an issue again. A smaller value of M means fewer communication steps. However, the number of grid points grows as the value of

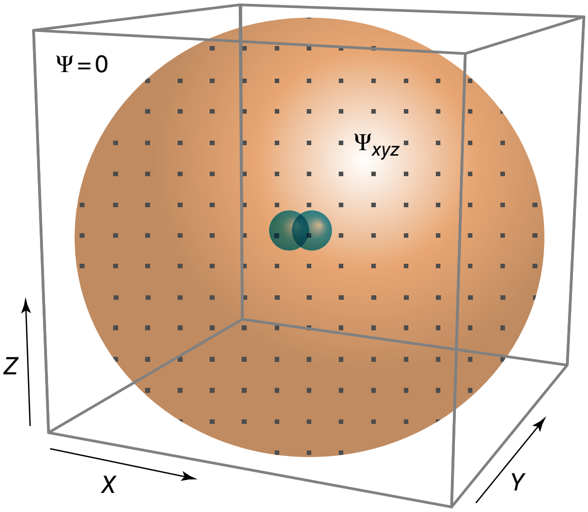


Figure 9.1 Example of a domain for finite differencing. Outside a sphere, the wave function is set to zero. The finite difference solution yields the value of the wave function at a grid point labeled by (x_i, y_j, z_k).

h^3, so it is often best to choose a modest size value for M and take a coarser grid spacing.

Nonlocality in ionic pseudopotential, $V_{\text{ion}}^{\text{p}}$, deserves special mention. Pseudopotentials come at a price. We omit superfluous degrees of freedom associated with core states when we use pseudopotentials, but in doing so we introduce a nonlocal operator, i.e. the potential at a point (x_i, y_j, z_k) depends on the value of the pseudopotential at neighboring points. Recall from Chapter 6 that nonlocality in the pseudopotential can be expressed in real space as an operator.

Nonlocality in the pseudopotential is localized to the core region and composes a very small fraction of the nonzero matrix elements in the Hamiltonian, but presents "communication" issues in handling the matrix operations. The additional computational load is more than compensated by the advantages of the pseudopotential for reasons pointed out earlier – the reduction in the number of states to only the valence states, which fixes the energy and length scales to the most weakly bound and extended states.

9.2.1 Special Diagonalization Methods: Subspace Filtering

A traditional method for finding eigenvalues finds the lowest one, then the next one, and so on. However, it is not necessary to proceed in such a manner. In the self-consistent loop (see Figure 7.3), our goal is to find the ground state density. The pathway we take to achieve this goal need not entail an eigenvalue by eigenvalue algorithm.

Consider a set of wave functions or orbitals that we know to be the correct solutions of the Kohn–Sham equation:

$$\Phi = \begin{bmatrix} \psi_1(r_1) & \psi_2(r_1) & \cdots & \cdots & \psi_m(r_1) \\ \psi_1(r_2) & \psi_2(r_2) & \cdots & \cdots & \psi_m(r_2) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \psi_1(r_N) & \psi_2(r_N) & \cdots & \cdots & \psi_m(r_N) \end{bmatrix}$$

where the $\psi_j(r_k)$ represents the Kohn–Sham orbital for the j-th eigenvalue at the grid point r_k. We assume there are m orbitals spanning N grid points. We can form a "density matrix," P, by taking $P = \Phi\Phi^T$ where Φ^T is the transpose of Φ. As before, we are going to assume that the matrix elements are real. The size of the diagonal of the matrix P is given by the number of grid points, r_i, so that each diagonal element gives the charge density, $\rho(r_i)$, at the grid point in question r_i is given by $\rho(r_i) = \psi_1(r_i)^2 + \psi_2(r_i)^2 \ldots + \psi_m(r_i)^2$. To get the diagonal elements, we need not have the matrix Φ; any rotation of this matrix will do. Suppose we construct a matrix U where $UU^T = I$, with I being the identity matrix. Then we could write

$$P = \Phi\Phi^T = \Phi UU^T\Phi^T = (\Phi U)(U^T\phi^T) = (\Phi U)(\Phi U)^T \tag{9.3}$$

where we use the matrix identity: $(AB)^T = B^T A^T$. In short, either Φ or ΦU will serve our purpose.

Can we devise an efficient way of getting ΦU? Suppose we consider a Hamiltonian of the form:

$$\mathscr{P}(H) = \sum_{n=1}^{M} a_n H^n \tag{9.4}$$

That is, we express the Hamiltonian in terms of a polynomial up to order M. We note that $H\psi_j(r) = E_j\psi_j(r)$ and consider a general function of the form:

$$\psi = \sum_{n=1}^{m} c_n \psi_n \tag{9.5}$$

This assumes that we can express a physically meaningful wave function by summing over the set $\{\psi_n\}$. We generate a new wave function by performing the following operation:

$$\hat{\psi} = \mathscr{P}(H)\psi = \mathscr{P}(H) \sum_{n=1}^{m} c_n \psi_n = \sum_{n=1}^{m} \mathscr{P}(E_n) c_n \psi_n \tag{9.6}$$

The wave function $\hat{\psi}$ will contain a sum over the m orbitals. The sum also contains a factor $\mathscr{P}(E_n)$. Suppose this factor is zero for any eigenvalue greater than the Fermi energy, E_f. If this is the case,

$$\hat{\psi} = \sum_{\mathrm{occup}} b_n \psi_n \tag{9.7}$$

We created a wave function that contains only components for the occupied orbitals. We "filtered out" components that contain empty states. If we continue to perform such an operation, we effectively find states that correspond to $\{\Phi U\}$, from which we can extract the self-consistent charge density.

A key aspect of filtering is to construct a polynomial, $\mathscr{P}$, that enhances the components in an energy region of interest and suppresses them elsewhere. An example is given in Figure 9.2. To the uninitiated, there are some "arcane" issues with this process. One issue is where to place the filter. If the eigenvalue spectrum is unknown, the energetics of the filter is unknown.

There are several ways to address this issue. Often, the spectrum is approximately known, e.g. the band width of a silicon nanostructure is usually not too different from that of bulk silicon so that one can make a reasonable guess as to the distribution of eigenvalues. Another option is to use an approximate method to assess the energy spectrum. As the computation proceeds, the filter can be reset.

Another issue is the nature of the polynomial itself. Any polynomial that has the behavior shown in Figure 9.2 will do. One common approach is to use Chebyshev polynomials. These polynomials are defined over the interval $|x| \leq 1$ where $P_0(x) = 1, P_1(x) = x$. A recursion relation exists for these polynomials: $P_{n+1}(x) = 2xP_n(x) - P_{n-1}(x)$. The polynomials are simple and the recursion relation expedites creating higher order polynomials. These attributes are what we want: simplicity and ease of implementation. Details of this method can be found in the literature [5–8].

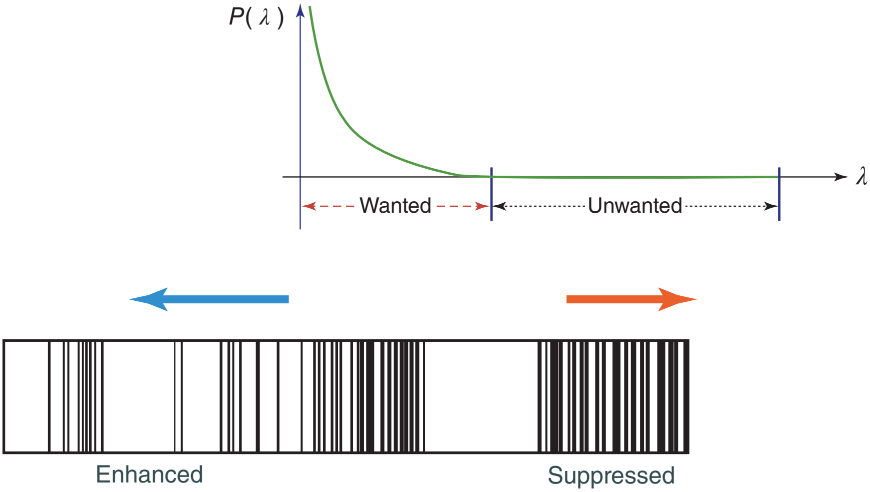


Figure 9.2 Schematic of a filtering polynomial. The polynomial $\mathscr{P}$ is large where we want to enhance the wave function components and small where we do not.

We can construct a filtering polynomial used in Figure 9.2 by an affine transformation of the Chebyshev polynomial. Filtering can dramatically speed up solutions to the Kohn–Sham equation. The self-consistent loop in Figure 7.3 is only slightly modified. Instead of a solution to the eigenvalue problem using standard diagonalization methods, the wave functions are filtered through a multiplication of the wave functions by a Chebyshev polynomial, $\mathscr{P}$. In Table 9.2, we give numbers. The speed up can be an order of magnitude faster compared to popular codes such as ARPACK, and there is no loss of accuracy.

We illustrate the use of this method by finding the eigenvalues for a large system. A convenient test case is a large nanocrystal of silicon. (We will discuss this system in more detail in Chapter 13.) We know the distribution of eigenvalues, the density of states, for crystalline silicon and can use this as a guide to whether our system is working. We take a fragment of bulk silicon and cap off any dangling bonds with hydrogen atoms. The resulting nanocrystal has a chemical formula of $Si_{20389}H_{3076}$. There are 84 632 valence electrons in this system. Without a pseudopotential, this system would contain 288 522. By any measure, at least using computers at the beginning of the twenty-first century, this is a very large number

Table 9.2 Run times to solve the Kohn–Sham problem for a nanocrystal of Si ($Si_{525}H_{276}$). ARPACK [9] and TRLan [10] are common public domain eigensolvers.

Method	Computation time (s)	Total nergy (ev/atom)
Chebyshev filtering	5947	−77.3168
ARPACK	62026	−77.3168
TRLan	26853	−77.3168

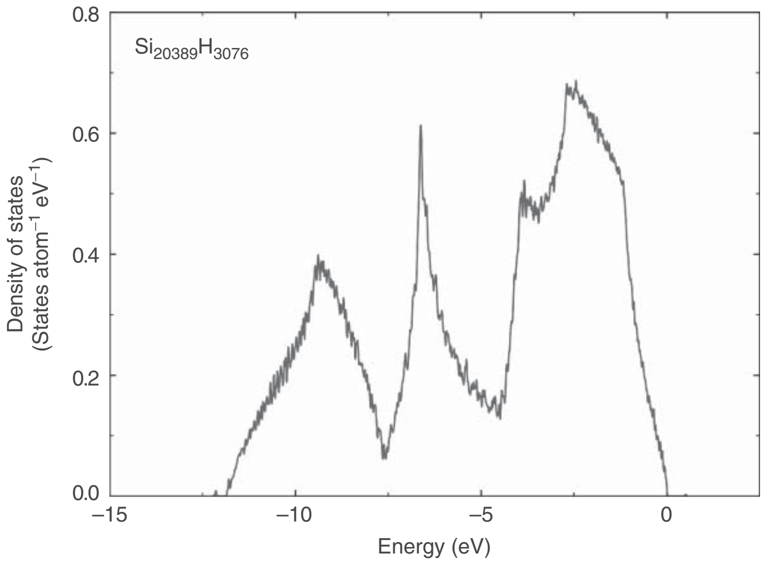


Figure 9.3 Density of states for a large nanocrystal of silicon.

of electrons. Putting two electrons in each state means we have 42 316 eigenvalues to solve.

It makes no sense to list such eigenvalues one by one, but we can make a histogram and illustrate the distribution of these states. The distribution of the states as a function of energy is effectively the density of states we discussed for a free electron gas (see, Eq. 4.9 and the related discussion). In Figure 9.3, the eigenvalues for the $Si_{20389}H_{3076}$ nanocrystal are displayed. We defer a detailed discussion of the results, save to make a few observations on the evolution of the electronic structure. The energy width of the occupied states spans ~12 eV. The top of the valence band is taken to be the zero energy reference. The gap between the highest occupied state and the lowest empty state is roughly 0.6 eV. These features agree with energy band computations for crystalline silicon. In general, the density of states of a nanocrystal assumes a crystalline silicon configuration with several thousand atoms. Obtaining a bulk-like density of states for large nanocrystals is a good test for any algorithm and the corresponding implementation. This calculation represents a "proof in principle" for large-scale computations.

References

1 Pulay, P. (1969). Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules. *Molecular Physics* 17: 197.

2 LeVeque, R.J. (2007). *Finite Difference Methods for Ordinary and Partial Differential Equations.* SIAM.

3 Reddy, J.N. (2006). *An Introduction to the Finite Element Method*, 3e. McGraw-Hill.

4 Fornberg, B. (1998). Calculation of weights in finite difference formulas. *SIAM Review* 40: 685.

5 Chelkowsky, J.R., Troullier, N., and Saad, Y. (1994). The finite-difference-pseudopotential method: electronic structure calculations without a basis. *Physical Review Letters* 72: 1240.

6 Kronik, L., Tiago, M.L., Alemany, M.M.G. et al. (2006). PARSEC - the pseudopotential algorithm for real-space electronic structure calculations: recent advances and novel applications to nano-structures. *Physica Status Solidi B* 243: 1063.

7 Zhou, Y., Saad, Y., Tiago, M.L., and Chelikowsky, J.R. (2006). Parallel self-consistent-field calculations via Chebyshev-filtered subspace acceleration. *Physical Review E* 74: 066704.

8 Schofield, G., Chelikowsky, J.R., and Saad, Y. (2012). A spectrum slicing method for the Kohn-Sham problem. *Computer Physics Communications* 183: 497.

9 Lehoucq, R.B., Sorensen, D.C., and Yang, C. (1998). *ARPACK USERS GUIDE: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods*. Philadelphia, PA: SIAM.

10 Wu, K. and Simon, H. (2000). Thick-restart Lanczos method for large symmetric eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications* 22: 602.

10

Atoms

> *A physicist is just an atom's way of looking at itself.*
>
> — Niels Bohr
>
> *If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis (or the atomic fact, or whatever you wish to call it) that all things are made of atoms …*
>
> — Richard Feynman

10.1 Energy Levels

The key to understanding the properties of an atom resides in the spatial and energetic distributions of the electrons that surround the nucleus. These distributions can be determined by a solution of the Kohn–Sham equation, which can quickly be solved for a spherical symmetric atom – as only one dimension is involved. The only limitation in this approach is the inherent accuracy or lack thereof in density functional theory.

In Figures 10.1 and 10.2, we show the Kohn–Sham energy levels for the outermost s and p states. We can use these levels to examine how the shells of an atom fill up. Qualitatively the s and p levels behave in a similar manner. As the shells fill, the energy levels monotonically decrease in energy until the shell is filled. When the next shell starts to fill, the energy levels experience a significant jump in energy. This shell structure reflects the orbital nature of the atomic wave functions. "Orbital-free" methods as in Thomas–Fermi theory do not produce shell structures, which is a serious deficiency.

The physical meaning of the Kohn–Sham levels is not particularly transparent. One might naively believe that the energy to remove an electron from an atom is the energy of the highest occupied energy level relative to the vacuum. Often this is not a bad estimate. However, within density functional theory, the levels cannot be directly associated with quantities related to observables, i.e. a quantity such

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

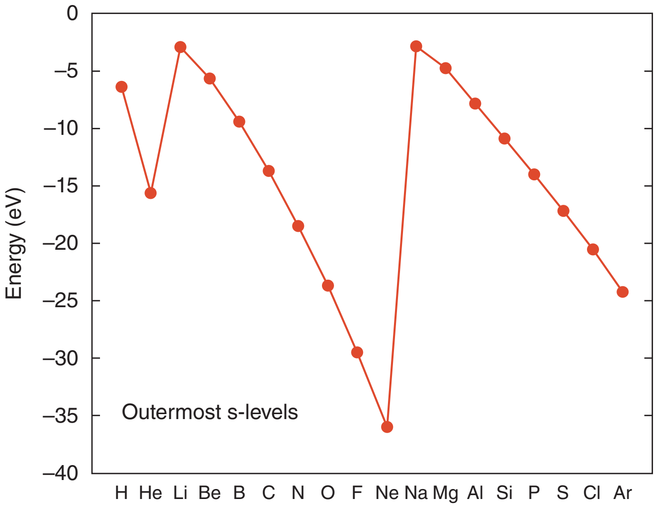


Figure 10.1 Energy levels for the s states for atoms in the first two rows of the periodic table.

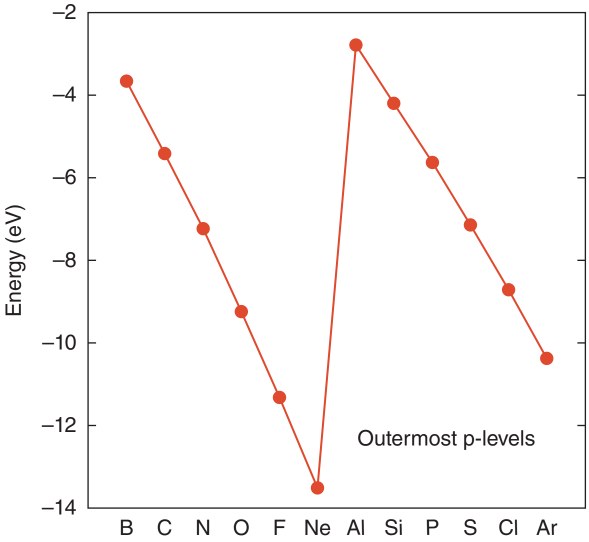


Figure 10.2 Energy levels for the p states for atoms in the first two rows of the periodic table.

as the energy to remove an electron. Nonetheless, the levels do follow the same trends expected for the energy involved in ionizing an atom.

10.2 Ionization Energies

The ionization energy of an atom may be determined by differencing the total energies of a neutral and an ionized atom:

$$I = E_{\text{total}}[n-1] - E_{\text{total}}[n] \tag{10.1}$$

where E_{total} is the total electronic energy for a neutral atom with n electrons. The sign convention is chosen so that the resulting ionization energy is positive. The second ionization energy can be determined from

$$I_2 = E_{\text{total}}[n-2] - E_{\text{total}}[n-1] \tag{10.2}$$

This is the energy to remove an electron from an ionized atom. The energy to remove two electrons is given by adding the first and second ionization energies: $I + I_2$

Within the framework of pseudopotential – density functional theory – H, Li, and Na are among the most problematic atoms when computing ionization energies. These atoms have only one valence electron as the corresponding atomic pseudopotential does not bind the inert core electrons.

When we form the Hartree and exchange-correlation potentials in density functional theory this valence electron may "screen itself," which is not a physically acceptable result. When we discussed the Hartree–Fock method, we noted that the self-interaction energy explicitly cancelled when combining the exchange and Hartree potentials. However, in density functional theory, as implemented in its simplest version, this might not happen and self-interaction or self-energy errors are created.

For one electron, we could circumvent this problem by ignoring the interaction of the electron with itself, i.e. we could take the unscreened ionic pseudopotential and not bother forming a Hartree and exchange potential (or an exchange-correlation potential). There is some logic to doing that, but this procedure is not the way to go. We need to be consistent in dealing with many-electron problems. We do not want to treat some atoms following one rule and others following a different rule. We mentioned this when we used density functional theory for estimating the ionization energy of the helium atom (see Chapter 7). There, we found that cancellation of errors resulted in a more accurate answer. In short, although approximations to the "true" density functional may not completely account for the self-interaction, the problem is often mitigated by error cancellation.

The first ionization energies are presented in Figure 10.3, and the ionization energies to remove two electrons are depicted in Figure 10.4. The calculations shown here were done using pseudopotentials within the local density approximation. These are simple computations that can be done with the computing power of a smart calculator, more or less. It is surprising how well theory and experiment agree, if we are willing to accept errors that are typically less than a few percent of the ionization energies.

We note one exception to the overall accuracy. Density functional theory predicts a monotonic increase in the ionization energy as the p-shell is filled. This does *not* agree with experiment, i.e. the ionization energy for the N–O–F sequence shows a minimum at O (this is also true to a lesser extent for P–S–Cl). We should not be surprised by this exception. In our quest for simplicity, we assumed "spinless" electrons. This assumption is not a good one when we have open shells as discussed in the following section.

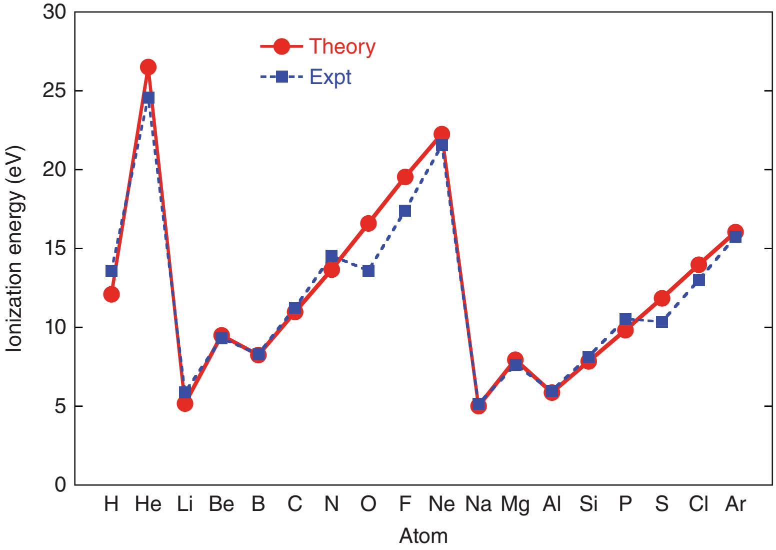


Figure 10.3 Ionization energies for atoms. Experiment is from the database at NIST (http://physics.nist.gov/PhysRefData/ASD/ionEnergy.html).

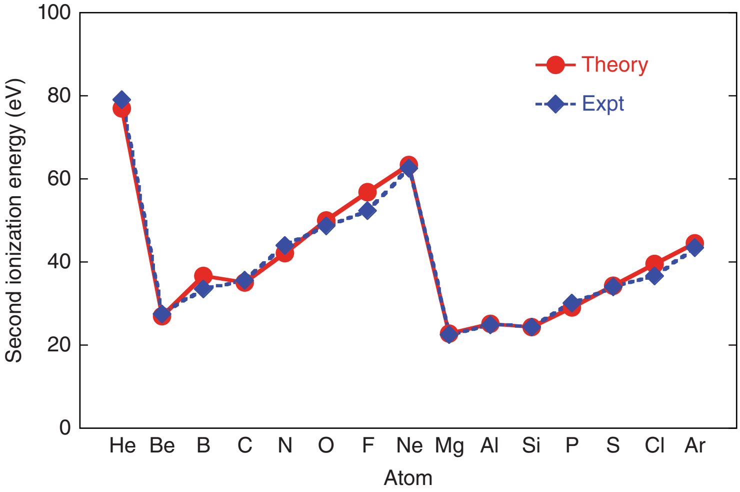


Figure 10.4 Energy to doubly ionize an atom. Experiment is from the database at NIST (http://physics.nist.gov/PhysRefData/ASD/ionEnergy.html).

10.3 Hund's Rules

Spin interactions can lower the energy of an atomic state. We need to consider such interactions. We describe some "rules" named after a German physicist – Friedrich Hund [1]. Hund recognized that angular moment is associated with the spin of the electron and with the motion of the electron. As a simple example, consider Bohr's model for the hydrogen atom. An electron moving in a circular path around the proton has angular momentum associated with its orbital motion. Classically, the angular momentum is given by $L = mvr$, which in the Bohr model is $L = (n\hbar/r)r = n\hbar$ from Eq. (2.4). The orbital momentum can be characterized by a vector, $\vec{L}$. An electron also carries spin and this contributes to the angular momentum too. We denote the spin angular momentum by $\vec{S}$. The total angular momentum is given by $\vec{J} = \vec{L} + \vec{S}$. For many-electron atoms, determining the possible angular momentum components can be messy, but the rules given by Hund are not.

Hund's rules for the ground-state configuration for a given shell of an atom are as follows.

- First, occupy the orbitals so that the total spin, S, is maximized.
- Second, maximize the value of the orbital angular momentum, L, consistent with the value of S.
- Third, the value of the total angular momentum, J, is $|L - S|$ when the shell is less than half full and equal to $L + S$ when the shell is more than half full.

Implicit in these rules are that one must observe the Pauli exclusion principle once Hund's rules have been executed. We can express the resulting ground state in "term symbol" notation. The term symbol is given by

$$^{2S+1}L_J$$

We list the term symbols in Table 10.1 for the first two rows of the periodic table using Hund's rules.

We consider some examples to make clear how Hund's rules work. Consider the carbon atom. The outer shell has a configuration of $2s^22p^2$. The 2s orbital is filled and need not enter our consideration. The maximum value of S occurs when we place the 2p electrons in different p states, e.g. p_x and p_z, with parallel spins. In this case $S = \sum_n s_n$=1 in units of $\hbar$. As is customary, we do not explicitly carry the $\hbar$ symbol.

Next, we deal with L. The maximum value of L occurs when we put one electron in an $m_l = 1$ and another in $m_l = 0$: $L = \sum_n (m_l)_n = 1$. For J, we recognize that the shell is less than half full so we take $J = |L - S| = 0$.

We now have sufficient information to label the ground-state term symbol; we characterize the total angular momentum $L = 0, 1, 2, \ldots$ by the letters $S, P, D, \ldots$. It is unfortunate that the total spin is also labeled by S, but we are stuck with it. In the example for carbon, $2S + 1$=3, $L = 1$, and $J = 0$, which results in the following notation for carbon:

$$^3P_0$$

Another example to be considered is the nitrogen atom. The outer shell has the configuration $2s^22p^3$. The maximum value of S is to fill the p_x, p_y, p_z orbitals with one electron each. This gives $S = 3/2$. The orbital momentum, L, vanishes as $L = \sum_n (m_l)_n = 0$, which leaves $J = S = 3/2$, and the term symbol is

$$^4S_{3/2}$$

The reader may complain, and rightfully so, that we are just labeling term values according to the lowest energy state, but we are not computing the energies involved. While it is possible to account for the different spectroscopic terms using density functional theory, we will not pursue this here. It is just too complicated and beyond the scope of this book.

Table 10.1 Ground-state term values and ionization energies for the atoms in first two rows of the period table.

Element	Ground-state shell structure	Term symbol	Ionization energy (eV)
Hydrogen	$1s^1$	$^2S_{1/2}$	13.60
Helium	$1s^2$	$^1S_0$	24.59
Lithium	$1s^22s$	$^2S_{1/2}$	5.39
Beryllium	$1s^22s^2$	$^1S_0$	9.32
Boron	$1s^22s^22p$	$^2P_{1/2}$	8.30
Carbon	$1s^22s^22p^2$	$^3P_0$	11.26
Nitrogen	$1s^22s^22p^3$	$^4S_{3/2}$	14.53
Oxygen	$1s^22s^22p^4$	$^3P_2$	13.63
Fluorine	$1s^22s^22p^5$	$^2P_{3/2}$	17.42
Neon	$1s^22s^22p^6$	$^1S_0$	21.56
Sodium	[Ne] 3p	$^2S_{1/2}$	5.14
Magnesium	[Ne] $3s^2$	$^1S_0$	7.65
Aluminum	[Ne] $3s^23p$	$^2P_{1/2}$	5.99
Silicon	[Ne] $3s^23p^2$	$^3P_0$	8.15
Phosphorus	[Ne] $3s^23p^3$	$^4S_{3/2}$	10.49
Sulfur	[Ne] $3s^23p^4$	$^3P_2$	10.36
Chlorine	[Ne] $3s^23p^5$	$^2P_{3/2}$	12.97
Argon	[Ne] $3s^23p^6$	$^1S_0$	15.76

Why do Hund's rules work?

The first rule can be rationalized from the Pauli exclusion principle. The exclusion principle mandates that two electrons cannot possess the same set of quantum numbers. For an open shell, the quantum numbers for an electron can differ in spin. Consider a silicon atom; the outermost shell is the $3p^2$. We could put two electrons with opposite spins in a state with $l = 1, m_l = 0$, or we could put two electrons with parallel spins in two spatially different states $l = 1, m_l = -1$ and $l = 1, m_l = 0$. In the latter case, the electrons on average are in spatially distinct orbitals. Electrons in singly occupied orbitals are less effectively screened from the nuclear charge and the electron–nuclear attraction energy is increased. This explanation is different from the "traditional one." Tradition holds that electrons in distinct orbitals experience a reduced electron–electron coulomb repulsion as on average they are further apart. Accurate quantum calculations do not support this explanation [2].

The second rule can be understood from a really simple picture of electrons moving in classical orbits. In this picture, atoms with higher orbital angular momentum contain more electrons moving in the same direction. Specifically, consider two electrons at the opposite ends of a circular orbit. Assume that they are moving in the same direction at the same rate. Compare this to a similar circular orbit in which the two electrons move in opposite directions at the same rate. In the latter case, the net angular momentum vanishes and the electrons will often approach each other. In the former case, the angular momentum is larger and the separation between the electrons is constant and at a maximum for a given orbit. Since the higher angular momentum case reduces the electron–electron repulsion, it will possess a lower energy.

The third rule results from considering spin–orbit interactions. A classical orbiting electron can generate a magnetic moment just like a current-carrying circular loop. This orbitally generated moment can then interact with the magnetic moment associated with the spin of the electron. For a single electron the energy of this interaction is lowest when the electron spin is opposite to the orbital angular momentum. As we fill a shell with additional electrons, we are restricted to choose states where the angular momentum is opposite to the spin; the low energy pairs of (m_l, m_s) are progressively used up. By the exclusion principle, when the shell is more than half full the state of the lowest energy necessarily has the spin parallel to the orbital moment. In this case, the sign of the energy is reversed. So we want $J = L + S$ instead of $J = |L - S|$. For those familiar with solid-state physics, a shell missing one electron can be thought of as a filled shell containing a "hole" state wherein the hole has an opposite sign from the electron.

10.4 Excited State Energies and Optical Absorption

When atoms are placed in a radiative field, they can gain energy by absorbing a photon (or lose energy by emitting a photon). There are two general rules based on fundamental laws of physics that govern the absorption or emission of a photon. The rules are based on (i) *energy conservation* and (ii) *symmetry considerations*. We will use a simple, but essentially correct, model that is often presented in textbooks [3].

Consider an isolated atom in the absence of any external fields. We assume the atom has one valence electron. We examine the behavior of the corresponding wave function, which could be a function of space and time. The solution of the time-dependent Schrödinger equation is given by

$$\mathscr{H}\Psi = i\hbar \frac{\partial \Psi}{\partial t} \tag{10.3}$$

If $\mathscr{H}$ is time independent, label it as $\mathscr{H}_0$. A solution for Ψ given $\mathscr{H}_0$ is

$$\Psi = \exp(-i\omega_n)\phi_n \tag{10.4}$$

where ϕ_n is the time-independent solution:

$$\mathscr{H}_0\Psi_n = \mathscr{H}_0 \exp(-i\omega_n)\phi_n = i\hbar \frac{\partial \exp(-i\omega_n)\phi_n}{\partial t} = \hbar\omega_n \exp(-i\omega_n)\phi_n \tag{10.5}$$

which can be rewritten as

$$\mathscr{H}_0\phi_n = \hbar\omega_n\phi_n = E_n\phi_n \tag{10.6}$$

Simple enough. Note that the charge density from Eq. (10.4), i.e. $|\Psi|^2$, is independent of time. Only the phase changes with time and phase plays no role in the charge density. Again, it is the "weird" nature of quantum mechanics that distinguishes "observables" from "nonobservables."

How does a radiation field change the wave function as a function of time and space? We take the field to be a homogeneous electrical field. (Atoms can couple to magnetic fields, but the coupling is much less important.)

Assuming a homogeneous field is reasonable. The spatial variation of light is on a much larger length scale when compared to the size on an atom. The wavelength for visible light varies from 400 to 700 nm and is considerably larger than the size of an atom whose size might be 0.1 nm. We consider a uniform sinusoidal variation of the light, which we can always generalize to more complex wave forms by superposing different variations.

The potential an electron experiences in such a field is given by

$$\Phi(z,t) = -\mathscr{E}_0\{\exp(i\omega t) + \exp(-i\omega t)\}z \tag{10.7}$$

The time dependence of the potential is given by a frequency ω. The potential is real as $\{\exp(i\omega t) + \exp(-i\omega t)\} = 2\cos(\omega t)$. This potential corresponds to a uniform field of magnitude $\mathscr{E}_0$. The field is polarized in the z-direction as determined by the gradient operator:

$$\vec{\mathscr{E}} = -\nabla\Phi(z,t) = \mathscr{E}_0\{\exp(i\omega t) + \exp(-i\omega t)\}\hat{z} \tag{10.8}$$

We can replicate this field as a potential energy:

$$V(z,t) = -e\Phi(z,t) = e\mathscr{E}_0\{\exp(i\omega t) + \exp(-i\omega t)\}z \tag{10.9}$$

The total Hamiltonian is formed by adding the time-independent potential to the time-dependent one: $\mathscr{H} = \mathscr{H}_0 + V(z,t)$. The time-dependent Schrödinger equation can now be written as

$$\mathscr{H}\Psi = [\mathscr{H}_0 + V(z,t)]\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{10.10}$$

We will consider $V(z,t)$ to be a small perturbation to $\mathscr{H}_0$.

Suppose we do the following. We consider our single electron to be initially in the state ϕ_0 and ask – what is the possibility it absorbs a photon and makes a transition to an excited state? For simplicity, we assume there is an empty state, ϕ_1, and no other states nearby. To approximate the excited state, we write

$$\Psi = \phi_0\exp(-i\omega_0 t) + a(t)\phi_1\exp(-i\omega_1 t) \tag{10.11}$$

At $t = 0$, we take $a = 0$. At later times $a(t)$ is a measure of the transition probability that the atom is excited to the ϕ_1 state. Since the effect of the field is small, the change in the Hamiltonian and the change in the time-independent states should both be small. From Eq. (10.10), we have

$$\begin{aligned}&[\mathscr{H}_0 + V(z,t)][\phi_0\exp(-i\omega_0 t) + a(t)\phi_1\exp(-i\omega_1 t)] = \\ &\qquad i\hbar\frac{\partial}{\partial t}[\phi_0\exp(-i\omega_0 t) + a(t)\phi_1\exp(-i\omega_1 t)]\end{aligned} \tag{10.12}$$

At this point, we use a "standard operating procedure." We multiply both sides of Eq. (10.12) by ϕ_1^* and integrate over all space. We make use of the fact that $H_0\phi_n = E_n\phi_n$ and that the time-independent wave functions obey

$$\int \phi_i^* \phi_j \, \mathrm{d}^3 r = \delta_{ij} \tag{10.13}$$

We neglect the product of the small perturbing potential with the excited state coefficient. The term containing the product "$a(t)V(z,t)$" involves the product of two "smallish" quantities, which we set to zero. By neglecting such terms we apply first-order perturbation theory [3].

This yields

$$i\hbar \frac{\mathrm{d}a}{\mathrm{d}t} = \mathscr{E}_0 \{\exp(i\omega t) + \exp(-i\omega t)\} \exp(i\omega_{10} t)\, z_{10} \tag{10.14}$$

where $\omega_{10} = \omega_1 - \omega_0$ and we define z_{10} as

$$z_{10} \equiv \int \phi_1^* \, z \, \phi_0 \, \mathrm{d}^3 r \tag{10.15}$$

To solve for $a(t)$ we do the integral:

$$a(t) = \frac{\mathrm{e}}{i\hbar} \mathscr{E}_0 \, z_{10} \int_0^t \{\exp(i\omega t) + \exp(-i\omega t)\} \exp(i\omega_{10} t) \, \mathrm{d}t \tag{10.16}$$

which yields

$$a(t) = \mathrm{e} \; \mathscr{E}_0 \; z_{10} \left\{ \frac{1 - \exp(i(\omega + \omega_{10})t)}{\hbar(\omega + \omega_{10})} - \frac{1 - \exp(-i(\omega - \omega_{10})t)}{\hbar(\omega - \omega_{10})} \right\} \tag{10.17}$$

This determines the time-dependent wave function. Once we have the wave function, we evaluate the resulting time-dependent dipole:

$$\mathrm{e}\mathscr{Z}(\omega, t) = \mathrm{e} \int z \, |\Psi(\vec{r}, t)|^2 \, \mathrm{d}^3 r \tag{10.18}$$

If we keep only terms linear in $a(t)$, then

$$\mathrm{e}\mathscr{Z}(\omega, t) = \mathrm{e}^2 \; \hbar \; \{a(t) \, z_{01} \; \exp(-i\omega_{10} t) + a(t)^* \, z_{10} \; \exp(i\omega_{10} t)\} \tag{10.19}$$

Neglecting some time-dependent terms that are not commensurate with the external electric field (they average to zero anyway), we get the following:

$$\mathrm{e}\mathscr{Z}(\omega, t) = \frac{\mathrm{e}^2 \; |\mathrm{z}_{10}|^2}{\hbar} \left\{ \frac{2\omega_{10}}{\omega_{10}^2 - \omega^2} \right\} \mathscr{E}_0(\exp(i\omega t) + \exp(-i\omega t)) \tag{10.20}$$

where we have explicitly written the cosine as a sum of exponents. Our work here is done.

In our simple model, the applied electric field induces a dipole, which couples to the applied field. We will not go into the details of radiation theory, save to say that the largest interaction between the atom and the field occurs by an induced dipole. The size of the induced dipole depends on the magnitude of $|\mathrm{z}_{10}|^2$.

This is easy to interpret. If the excited state has the same symmetry as the initial state, then a linear combination of these states will not have any dipole components. (As an example, suppose ϕ_0 and ϕ_1 are both even functions such as the 1s and 2s states in hydrogen.) The electric field will not couple to these states, i.e. the atom will not absorb or emit dipole radiation between these states.

The other notable factor in Eq. (10.20) is the pole or singularity that exists when $\omega_{10} = \omega$. This pole indicates a resonance between the time dependence of the field and the response of the atom. A classic example of resonance occurs when one pumps energy into a spring by applying a time-dependent force that compresses and expands the spring. If the frequency of the applied force is commensurate with the characteristic vibration of the spring, resonance will occur. In the quantum case, resonance means that energy from the applied field can be absorbed by the atom. The resonance occurs when energy is conserved; the energy of photon equals the energy to promote an electron: $\hbar\omega = E_1 - E_0 = \hbar\omega_{10}$.

Our expression can be generalized to additional energy levels, corresponding to additional excited states, as follows:

$$\mathrm{e}\,\mathscr{Z}(\omega, t) = \sum_j \frac{\mathrm{e}^2\,|\mathrm{z}_{j0}|^2}{\hbar}\left\{\frac{2\omega_{j0}}{\omega_{j0}^2 - \omega^2}\right\}\mathscr{E}_0(\exp(i\omega t) + \exp(-i\omega t)) \tag{10.21}$$

This is a straightforward procedure. We add excited states to Eq. (10.11) and "repeat" the derivation.

We can define an "oscillator strength," which is a measure of the transition probability, as follows:

$$f_j \equiv \frac{2m}{\hbar^2}\,\hbar\omega_{j0}|z_{10}|^2 \tag{10.22}$$

It is easy to show that the oscillator strengths obey the so-called Thomas–Reiche–Kuhn sum rule [3]:

$$\sum_j f_j = 1 \tag{10.23}$$

We can use the definition of the oscillator strength to write

$$\mathrm{e}\,\mathscr{Z}(\omega, t) = \frac{\mathrm{e}^2}{m}\sum_j\left\{\frac{f_i}{\omega_{j0}^2 - \omega^2}\right\}\mathscr{E}_0(\exp(i\omega t) + \exp(-i\omega t)) \tag{10.24}$$

This expression can be recast by defining an atomic *polarizability*, $\alpha(\omega)$, so that

$$\mathrm{e}\,\mathscr{Z}(\omega, t) = \alpha(\omega)\,\mathscr{E}_0(\exp(i\omega t) + \exp(-i\omega t)) = \alpha(\omega)\mathscr{E}(\omega, t) \tag{10.25}$$

where

$$\alpha(\omega) = \frac{\mathrm{e}^2}{m}\sum_j \frac{f_i}{\omega_{j0}^2 - \omega^2} \tag{10.26}$$

This quantity defines how the charge density of the atom is altered by an electric field and is called a "response function." Since we only include linear powers of the electric field, this is also called a "linear response" function.

There are some subtle issues with Eq. (10.26). We should always worry about a pole in a physical quantity. Rarely do physical quantities diverge. Textbooks often invoke *ad hoc* explanations such as defects or impurities, friction or heat loss, or some other factors that we left out to explain why we do not get a real divergence [4].

There is a clear, but maybe not obvious, reason for existence of the pole. We recall that the perturbing field is of the form:

$$\mathscr{E} = \mathscr{E}_o \cos(\omega t)$$

Think about time in this expression. When was the field applied? There is no beginning or end to the field. The field exists unchanged a long, long time ago and unchanged far, far into the distant future. Physically, this means that the sinusoidal behavior of the field is known with "infinite" precision as one could measure it for infinite time. Clearly, this situation is unphysical compared to an experiment done over some finite time span.

Specifically, what happens if the field is slowly turned on? Suppose the sinusoidal function turns on such that $\mathscr{E} = \mathscr{E}_t \exp(\gamma t) \exp(i\omega t)$. For $t \to -\infty$ the field would be zero. Suppose for $t > 0$ we take $\gamma = 0$. This field slowly turns on and then oscillates with a constant amplitude. The derivation we used would be similar. However, we add an imaginary part to the frequency: $\omega \to \omega + i\gamma$. We remove the pole by writing

$$\alpha(\omega) = \frac{\mathrm{e}^2}{m} \sum_j \frac{f_i}{\omega_{j0}^2 - \omega^2 - i\gamma\omega} \tag{10.27}$$

Our polarization function is now complex, but well behaved. We separate out the real and imaginary parts. We also convert the polarization function to a more commonly used "dielectric function." The dielectric function relates the applied electric field and the response of the system to the field. It is easily generalized to condensed matter systems.

The polarizability of an atom is related to the electric field by $p = \alpha\mathscr{E}$ where p is the induced dipole. The electric polarizability of a weakly interacting ensemble of atoms can be written as $P = N\alpha p = \chi E$ where χ is the electrical susceptibility. The combination of the applied field and the polarizability is given by an elexctric displacement vector: $D = E + 4\pi P = (1 + 4\pi\alpha)E = \varepsilon E$. ε is the dielectric function, which we often want to know. It can be related to a variety of optical properties such as a complex index of refraction and the reflectivity of an object.

The real and imaginary parts of the dielectric function are often denoted as

$$\varepsilon = \varepsilon_1 + i\varepsilon_2$$

Using this notation, we have

$$\varepsilon_1(\omega) = 1 + \sum_j \frac{f_i\, \omega_\mathrm{p}^2(\omega_{j0}^2 - \omega^2)}{(\omega_{j0}^2 - \omega^2)^2 + \gamma^2\omega^2} \tag{10.28}$$

and

$$\varepsilon_2(\omega) = \sum_j \frac{f_i\, \omega_\mathrm{p}^2 \gamma\omega}{(\omega_{j0}^2 - \omega^2)^2 + \gamma^2\omega^2} \tag{10.29}$$

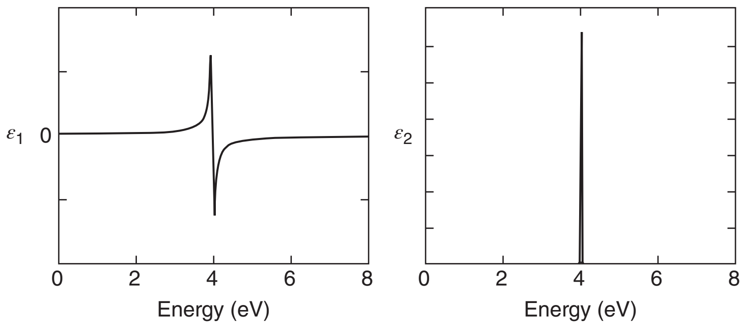


Figure 10.5 Schematic of dielectric response functions. A sharp resonance occurs at 4 eV.

where we have defined a new frequency, the plasma frequency: $\omega_\mathrm{p}^2 = 4\pi N\mathrm{e}^2/m$, where N is the number of electrons per unit volume. The plasma frequency corresponds to a resonant frequency of uniform electron density when displaced from a uniform positive background of the same charge density.

With the notable exception of the oscillator strength in Eq. (10.28), this expression is what we would expect for a damped harmonic oscillator. Some elementary books [4] introduce optical excitations by considering an electron attached to an atom by a spring in a viscous medium (to account for the damping.)

In Figure 10.5, we illustrate the essential features of a dielectric function for an ensemble of atoms where one transition dominates the spectrum. The resonant behavior illustrated is quite common for many spectral features. It qualitatively resembles the absorption in a simple semiconductor such as silicon or germanium, although the width of the absorption line is wrong because in solids there are energy bands and not energy levels as in an atom. The width of the imaginary dielectric function is dependent on γ. In the figure, γ is taken to be small. In the limit of $\gamma \to 0$, the width of the absorption line becomes vanishingly small and the line corresponds to a "delta function." A small line width would imply a long measurement of the absorption process, at least with respect to $1/\gamma$.

Example 10.1 ***Time Evolution of Atomic Wave Functions*** We just discussed the interaction of light with an atom by using perturbation theory. We found the probability of an atom making an electronic transition by absorbing light using the unperturbed wave functions. There is a more direct way of solving the problem by directly computing the time evolution of a wave function in the presence of an electric field. We illustrate the essential features of this approach for a sodium atom.

We can write down a Kohn–Sham time-dependent equation:

$$\mathscr{H}\phi_i = \frac{-\hbar^2\nabla^2\phi_i}{2m} + V_\mathrm{ext}\phi_i + V_\mathrm{H}\phi_i + V_\mathrm{xc}\phi_i = i\hbar\frac{\partial}{\partial t}\phi_i \tag{10.30}$$

where the external potential can be time dependent, e.g. it could include the ionic pseudopotential and an external field. This equation resembles what one might have guessed for a time-dependent Schrödinger equation in density functional theory.

This simplicity is deceiving! The equation has a number of assumptions in it. Specifically, the derivation of this equation assumes that the response of the system is local both in "time" and in "space." The history of the system does not enter our discussion as the system is local in time. The behavior of one part of the system interaction with the field does not affect other parts of the system as the system is local in space. It is hard to decide without doing a computation if the assumption of spatial and temporal locality is a bad one.

We dealt with locality in space in earlier chapters. For example, the *local* density approximation assumes that the spatial dependence of the exchange-correlation potential depends only on the point in question. For many systems, this is all right and works pretty well. The locality of time means that the potential responds instantaneously to any change in an applied field. This assumption is more complicated, both conceptually and computationally. We will need to test it. With these assumptions, the only explicit time dependence resides within the charge density. We assume the functional itself is not a function of time: $V_{\text{xc}} = V_{\text{xc}}[\rho(t)]$.

We can solve Eq. (10.30) in a formal sense by writing

$$\phi_i(r,t) = \exp\left\{\frac{1}{i\hbar}\int_0^t \mathscr{H}(r,t')\mathrm{d}t'\right\}\phi_i(r,0) \tag{10.31}$$

This an odd equation as $\mathscr{H}$, which is an operator, resides in an exponential. The equation has meaning only in terms of a Taylor series. Suppose we consider a short time interval, Δt, and we know the solution at $t = 0$. We assume the Hamiltonian $\mathscr{H}$ does not change appreciably over this interval. A solution for ϕ_i can then be written as follows:

$$\phi(r,t+\Delta t) = \sum_{k=0}^{\infty}\frac{1}{k!}\left[\frac{\Delta t H}{i\hbar}\right]^k \phi_i(r,t) \tag{10.32}$$

Clearly, this solution is correct in the limit of $\Delta t = 0$. This relation allows us to "step" the equation out by first computing $\phi(r,\Delta t)$ from $\phi_i(r,0)$ and then repeating, e.g. knowing $\phi(r,\Delta t)$ we can find $\phi(r,2\Delta t)$.

We now consider an isolated sodium atom. Suppose we know the solution for this atom with no applied field: $\mathscr{H}^0\phi_0^0(r) = E_0\phi_0^0(r)$. The various zero superscripts mean we are considering an atom in the absence of any applied fields. Since sodium has only one valence electron, we label the ground state also by a zero subscript. Also, we are going to assume that the Hamiltonian does not change in time as the ion core is unaffected by the applied field and for this particular example we will not screen the valence electron by its own charge.

Next, we apply a uniform electric field, $\mathscr{E}$, to the atom. We do this for an infinitesimal amount of time. As an analogy, consider striking a bell with a hammer. The bell will ring in respond to the strike and the audio signal can be broken down into some fundamental resonances of the bell. Here, we hit the atom with this uniform field and then analyze how the atom "rings."

Suppose we take

$$V_{\text{ext}}(r,t) = -\mathrm{e}\mathscr{E}z\delta(t) \tag{10.33}$$

At $t = 0$ we write our Hamiltonian as

$$\mathscr{H}(t = 0) = \mathscr{H}^0(t = 0) - \frac{e\mathscr{E}z}{\Delta t} \tag{10.34}$$

where the delta function is replaced by $1/\Delta t$. Our initial wave function has the form:

$$\phi_0(r, t = 0^+) = \exp(ikz)\phi_0^0(r) \tag{10.35}$$

This is not hard to see. Consider Eq. (10.31). The integral over the delta function integral yields

$$\int_{0^-}^{o^+} \mathscr{H}(r, t')\mathrm{d}t' = -e\mathscr{E}z \tag{10.36}$$

which gives $k = e\mathscr{E}z/\hbar$ in Eq. (10.35). It is relatively easy to propagate the initial state in time:

$$\phi_0(r, \Delta t) = (1 + i\mathscr{H}^0\Delta t/\hbar - (\mathscr{H}^0\Delta t/\hbar)^2/2 + \cdots)\ \exp(ikz)\phi_0^0(r) \tag{10.37}$$

We know the solution ϕ_0^0 without the field and the Hamiltonian, $\mathscr{H}^0$, so this operation is straightforward. The propagation step can be accelerated by taking more terms in the expansion or by using different algorithms for stepping the wave function. Of course, either way the problem is more difficult. Often simplicity is a better option. We let the computer handle the increased workload.

If the wave function is known as a function of time, then the charge density is also known as a function of time. This is easy for a one-electron atom such as sodium. We just square the occupied orbital:

$$\rho(r, t) = -e|\phi_0(r, t)|^2 \tag{10.38}$$

The next step is to find the dipole induced by the electric field. This follows from the charge density:

$$z(t) = \int z\ \rho(r, t)\ \mathrm{d}^3r \tag{10.39}$$

In Figure 10.6, we plot the induced dipole as a function of time. A harmonic oscillation is present, but what is the frequency of this oscillation? We find the

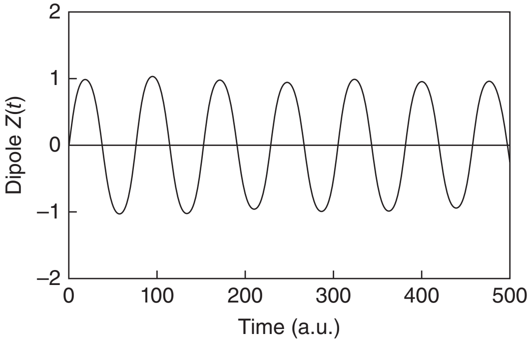


Figure 10.6 Time dependence of an induced dipole for a sodium atom that is pulsed by a uniform electric field.

frequency as previously. We compute a power spectrum by taking an integral of the form:

$$I(\omega) \propto \left| \int_0^\infty z(t)\ \exp(-i\omega t)\ \mathrm{d}t \right|^2 \tag{10.40}$$

We expect $I(\omega)$ to be a delta function response for a dipole with a pure harmonic oscillation. Performing the integral presents a few obstacles. For starters, does one really have to integrate over all time? The answer depends on the desired resolution. The longer we can observe the dipole, the better the resolution. We have previously addressed this problem. Suppose we assume a pure harmonic response: $z(t) = z_0 \exp(-i\omega_0 t)$. Then, assuming a proportionality constant, I_0, we write

$$I(\omega) = I_0 \left| \int_0^\infty \exp(-i(\omega - \omega_0)t)\ \mathrm{d}t \right|^2 \tag{10.41}$$

We easily evaluate this integral, save for the limit at infinity. We address this limit by assuming a damping term: $z(t) = z_0 \exp(-\gamma t) \exp(-i\omega_0 t)$ and consider $\gamma \to 0$. Mathematically this is not proper, but the physics is all right. Now our integral becomes

$$I(\omega) = I_0 \left| \int_0^\infty \exp(-i(\omega - \omega_0 + i\gamma)t)\ \mathrm{d}t \right|^2 \tag{10.42}$$

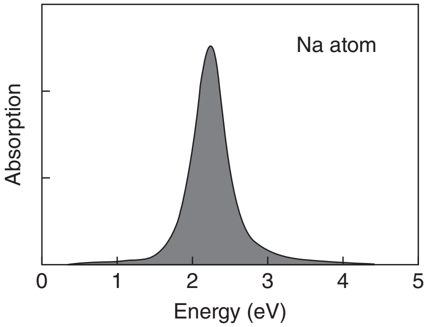
which yields

$$I(\omega) = \frac{I_0}{(\omega - \omega_o)^2 + \gamma^2} \tag{10.43}$$

This functional form of $I(\omega)$ is well known and is called a "Lorentzian" function or line shape. In Figure 10.7, we illustrate a Lorentzian function as determined from the calculated dipole in Figure 10.6. This plot uses a fairly large value of γ so the width is much larger than the experimental one as shown in Figure 10.8. The experimental peak occurs at about 590 nm, or about 2.1 eV. This value agrees with our computation to within a few percent.

The approach we illustrated for the sodium atom works well, but it is highly simplified. In a real system, as the wave functions change with time, we need to update the Hartree and exchange-correlation potentials. For sodium, we did not screen the electron. So these potentials were taken to be zero. Our Hamiltonian,

Figure 10.7 Power spectrum corresponding to the absorption spectrum for a sodium atom.



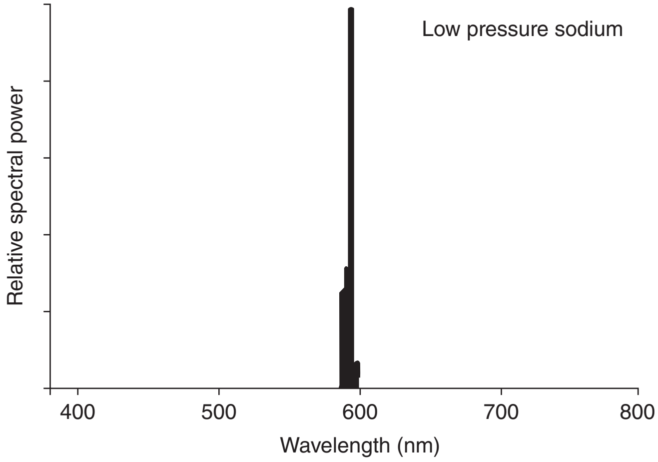


Figure 10.8 Experimental absorption spectrum from https://commons.wikimedia.org/wiki/File:Low-pressure_sodium_lamp_spectrum.svg.

$\mathscr{H}_0$, contains only the ionic pseudopotential, which is time independent as it did not change with the valence wave function. In a real system, $\mathscr{H}_0$ is updated after each time step reflecting changes in the charge density with time. The computations can be time consuming.

The theory illustrated in our example is called "time-dependent density functional theory," abbreviated as TDDFT. TDDFT works pretty well for simple systems such as a sodium atom and small metal clusters. However, the "local" approximations in time and space assumed in TDDFT may not be very accurate for all systems. Such approximations are the topics of current research [5, 6].

10.5 Polarizability

Here, we consider some general procedures for computing induced dipoles in the static limit. We could obtain this quantity by taking the static limit of Eq. (10.27) where we generalized the case to include all possible transitions, n:

$$\alpha(\omega = 0) = \frac{e^2}{m} \sum_n \frac{f_n}{\omega_n^2} \tag{10.44}$$

This is a useful expression that can readily be computed, provided one does not need many empty states to converge the sum. The number of states required is related to the distortion in the ground-state charge density as a function of the applied field. If there is a large distortion in the density, then more states are needed to replicate the changes in the ground-state density. This can be a limiting factor in any computations.

A more general way of computing this quantity is to include the static field directly in the Hamiltonian:

$$\left[\frac{-\hbar^2\nabla^2}{2m} + V_{\text{eff}}(\vec{r}) - e\vec{\mathscr{E}} \cdot \vec{r}\right] \psi(\vec{r})_i = E_i \psi(\vec{r})_i \tag{10.45}$$

where V_{eff} contains the ionic pseudopotential, the Hartree potential, and the exchange-correlation potential, and $\mathscr{E}$ is the uniform electric field. To determine the polarizability, we solve Eq. (10.45) in the presence of the electric field and find the induced dipole, $\vec{\mu}$, given by

$$\vec{\mu} = \int \vec{r}\ \rho(\vec{r})\ \mathrm{d}\vec{r}$$

In the case of a spherically symmetric atom the polarizability is independent of direction and the average dipole is a constant: $\langle \vec{\mu} \rangle = \mu$. The polarizability is given by $\alpha = \mathrm{d}\mu/\mathrm{d}\mathscr{E}$ where the derivative can be computed using finite differencing.

Does the result depend on the size of the applied field? Well it could, but in practice it need not. In the limit of a vanishingly small field, there can be numerical issues, i.e. the wave functions with and without field cannot be accurately computed. In the opposite limit, the field can be so large as to ionize the atom! Clearly, we need something within these limits. This can be determined by plotting the polarizability of the atom as a function of the applied field. Stable values for the polarizability can often be obtained, even when the applied field changes by an order of magnitude or more. Typical values for applied fields range from 10^{-4} to 10^{-3} a.u. If properly done, either method, perturbation theory or finite fields, should yield the same results. We will discuss this further in Chapter 11.

In Table 10.2, computations for the polarizability of noble gas atoms are given [8]. There are two notable trends. Theory predicts larger values than experiment and the relative difference decreases for the heavier atoms, i.e. the polarization for He is larger by about 20% whereas this difference is only about 5% for Xe. The source of this discrepancy is thought to arise from a flaw in the exchange-correlation potential as defined with density functional theory. If we remove an electron from an atom, we expect the potential at large distances to decay as $1/r$. Simple electrostatics predicts this. However, most functionals yield a potential that decays exponentially. Numerical studies show that *ad hoc* asymptotically corrected functionals can do much better [8].

Table 10.2 Static electric dipole polarizabilities of noble gas atoms calculated using the finite-field methods. Experiment is from Ref. [7].

	Polarizability (Å^3)	
Atom	**Theory**	**Experiment**
He	0.246	0.205
Ne	0.452	0.396
Ar	1.785	1.641
Kr	2.682	2.484
Xe	4.264	4.044

References

1 Herzberg, G. (1944). *Atomic Spectra and Atomic Structure*. Dover Publications.

2 Levine, I.M. (1991). *Quantum Chemistry*, 4e. Prentice-Hall.

3 Ziman, J.M. (1979). *Principles of the Theory of Solids*, 2e. Cambridge University Press.

4 Hummel, R.E. (2013). *Electronic Properties of Materials*, 4e. Springer-Verlag.

5 Marques, M.A.L., Ullrich, C.A., Nogueira, F. et al. (eds.) (2006). *Time-Dependent Density Functional Theory*. Springer-Verlag.

6 Ullrich, C. (2012). *Time-Dependent Density-Functional Theory: Concepts and Applications*. Oxford University Press.

7 Miller, T.M. and Bederson, B. (1988). Electric-dipole polarizability measurements. *Advances in Atomic and Molecular Physics* 25: 37–60.

8 Vasiliev, I. and Chelikowsky, J.R. (2010). Real-space calculations of atomic and molecular polarizabilities using asymptotically correct exchange-correlation potentials. *Physical Review A* 82: 012502.

11

Molecules

> *Life … is a relationship between molecules.*
>
> —Linus Pauling

11.1 Interacting Atoms

In Section 8.1 we focused on numerical approaches to interacting atoms, and discussed how quantum mechanics could account for the binding of two hydrogen atoms. We employed a theory by Heitler and London. The overall success of their theory is a significant result in understanding the chemical bond. Suppose their theory had failed? Such a failure would cast suspicions on quantum theory for all chemical interactions. Of course, that did not happen and we need not deal with alternate histories. Rather, we are confident that quantum theory works.

Molecules represent stable interactions between atoms. In contrast to "atomic clusters," which are also made from interacting atoms, molecules can be condensed into solids where the molecular unit has meaning. The simplest molecules are diatomics, such as H_2, O_2, Cl_2, N_2, NO, and CO.

Some of these diatomics are notably important to our very existence. The diatomic gases O_2 and N_2 make up over 99% of the earth's atmosphere. Understanding the properties of these elemental gases tells us much about the atmospheric ocean that shelters us from the vacuum of space.

We begin by focusing on the properties of several of these diatomics. We indicate how quantum theory can be used to compute the binding energy, the equilibrium bond length, and vibrational modes of diatomics. We will also consider interactions of the molecules with light and electric fields. We consider some more complex molecules for show.

11.2 Molecular Orbitals: Simplified

An orbital description of diatomic molecules is a good starting point. We label the two atoms by "A" and "B." The wave function for the diatomic are given as

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

follows:

$$\Psi_{\mathrm{AB}} = \sum_{i,j} \alpha_{i,j}\ \phi_j^i \tag{11.1}$$

The sum i runs over the *types of atoms*; in this case, $i = \mathrm{A}$ or B. The sum j runs over the *orbital character*, e.g. $j = \mathrm{s, p, d}, \ldots$.

We can be more specific. Suppose we consider a C_2 molecule, we might write

$$\begin{aligned}\Psi_{\mathrm{C}_2} &= \alpha_{\mathrm{A,s}}\phi_{\mathrm{s}}^{\mathrm{A}} + \alpha_{\mathrm{A,p}_x}\phi_{\mathrm{p}_x}^{\mathrm{A}} + \alpha_{\mathrm{A,p}_y}\phi_{\mathrm{p}_y}^{\mathrm{A}} + \alpha_{\mathrm{A,p}_z}\phi_{\mathrm{p}_z}^{\mathrm{A}} \\ &\quad +\alpha_{\mathrm{B,s}}\phi_{\mathrm{s}}^{\mathrm{B}} + \alpha_{\mathrm{B,p}_x}\phi_{\mathrm{p}_x}^{\mathrm{B}} + \alpha_{2,\mathrm{p}_y}\phi_{\mathrm{p}_y}^{\mathrm{B}} + \alpha_{\mathrm{B,p}_z}\phi_{\mathrm{p}_z}^{\mathrm{B}}\end{aligned} \tag{11.2}$$

Some clarification is in order. There are two carbon atoms. We label one of the atoms A and the other B as they exist on different sites. We consider s, p_x, p_y, p_z orbitals on each atom. As one might guess, keeping track of the various indices for large systems means being a good bookkeeper.

As is "standard practice," we solve the following eigenvalue problem

$$\mathscr{H}\,\Psi_{\mathrm{C}_2} = E\ \Psi_{\mathrm{C}_2} \tag{11.3}$$

by multiplying both sides by $\phi_{j'}^{*i'}$ and integrating over all space. This procedure yields a secular equation with matrix elements of the form

$$\int \phi_{j'}^{*i'}\mathscr{H}\ \phi_j^i\ \mathrm{d}^3r \quad \text{and} \quad \int \phi_{j'}^{*i'}\ \phi_j^i\ \mathrm{d}^3r \tag{11.4}$$

We consider an *orthonormal* set of basis functions:

$$\int \phi_{j'}^{*i'}\phi_j^i \mathrm{d}^3r = \delta_{ii',jj'} \tag{11.5}$$

When $i = i'$ and the orbitals are on the same site, there are two cases of merit. If the orbitals are simply atomic orbitals, which is a common procedure, the "usual rules apply." If $j = j'$ we take the orbitals to be normalized to unity. If $j \neq j'$, we take the orbitals to be orthogonal and the matrix element vanishes. When $i \neq i'$, the sites are different. We could assume the matrix element vanishes as the orbitals are not on the same atom and separated by a bond length. Clearly, this will be the case if the orbitals are strongly localized on each atom. Neglecting these matrix elements simplifies the problem enormously. If one does not make this approximation, there are "overlap" matrix elements between the orbitals. There are ways to handle such matrix elements, but for the sake of simplicity we will assume that the overlap matrix elements vanish.

Matrix elements involving the Hamiltonian can be handled in a similar manner. If the orbitals are on the same site $i = i'$, we have two cases. If j is an s-state and j' is an s state, then we have

$$\int \phi_{j'}^{*i}\mathscr{H}\phi_j^i\ \mathrm{d}^3r = E_{\mathrm{s}} \quad j = j' = \mathrm{s} \tag{11.6}$$

and if j is a p-state and j' is a similar p state, e.g. both are p_x states then,

$$\int \phi_{j'}^{*i}\mathscr{H}\phi_j^i\ \mathrm{d}^3r = E_{\mathrm{p}} \quad j = j' = \mathrm{p}_x, \mathrm{p}_y, \text{or } \mathrm{p}_z \tag{11.7}$$

Otherwise, this matrix element vanishes.

A complex situation occurs for $i \neq i'$, i.e. when one orbital is on atom A and the other orbital is on atom B. For simplicity, we wish to keep the most significant terms. What constitutes a "significant term?" It is hard to answer this without some experience. Fortunately, this experience exists.

Here is the story. An s orbital on atom A and an s orbital on atom B yield a significant matrix element:

$$\int (\phi_{\mathrm{s}}^{\mathrm{A}})^* \mathcal{H} \phi_{\mathrm{s}}^{\mathrm{B}} \, \mathrm{d}^3 r = \mathrm{s}^{\mathrm{A}}\mathrm{s}^{\mathrm{B}} \tag{11.8}$$

For s orbitals interacting with p, the matrix element can be neglected.

For the p states, we consider the following. Suppose the diatomic molecule is oriented along the x-axis. Then we take

$$\int (\phi_{\mathrm{p}_x^{\mathrm{A}}})^* \mathcal{H} \phi_x^{\mathrm{B}} \, \mathrm{d}^3 r = \mathrm{p}_x^{\mathrm{A}}\mathrm{p}_x^{\mathrm{B}} \tag{11.9}$$

We also allow p states in the direction perpendicular to interact, and we have this matrix element to consider:

$$\int \phi_{j'}^{*i'} \mathcal{H} \phi_j^{i} \, \mathrm{d}^3 r = \mathrm{p}_y\mathrm{p}_y'\mathrm{or} \ \ \mathrm{p}_z\mathrm{p}_z' \quad j = \mathrm{p}_y, \mathrm{p}_z \qquad j' = \mathrm{p}_y, \mathrm{p}_z \quad i \neq i' \tag{11.10}$$

The notation means that this integral is not zero if the orbitals at different sites are the same p states, e.g. p_y on site 1 with p_y on site 2. We neglect matrix elements where this is not the case, i.e. we neglect the matrix element with p_y on site A and p_z on site B. Quantum chemists often neglect such matrix elements and characterize the neglect of such elements as "zero differential overlap."

We organize our secular equation in a matrix format.

$$\begin{pmatrix} E_{\mathrm{s}} & \mathrm{s^A s^B} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathrm{s^A s^B} & E_{\mathrm{s}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & E_{\mathrm{p}} & \mathrm{p}_x^{\mathrm{A}}\mathrm{p}_x^{\mathrm{B}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{p}_x^{\mathrm{A}}\mathrm{p}_x^{\mathrm{B}} & E_{\mathrm{p}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & E_{\mathrm{p}} & \mathrm{p}_y^{\mathrm{A}}\mathrm{p}_y^{\mathrm{B}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{p}_y^{\mathrm{A}}\mathrm{p}_y^{\mathrm{B}} & E_{\mathrm{p}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & E_{\mathrm{p}} & \mathrm{p}_z^{\mathrm{A}}\mathrm{p}_z^{\mathrm{B}} \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathrm{p}_z^{\mathrm{A}}\mathrm{p}_z^{\mathrm{B}} & E_{\mathrm{p}} \end{pmatrix} \begin{pmatrix} \alpha_{\mathrm{s}}^{\mathrm{A}} \\ \alpha_{\mathrm{s}}^{\mathrm{B}} \\ \alpha_{\mathrm{p}_x}^{\mathrm{A}} \\ \alpha_{\mathrm{p}_x}^{\mathrm{B}} \\ \alpha_{\mathrm{p}_y}^{\mathrm{A}} \\ \alpha_{\mathrm{p}_y}^{\mathrm{B}} \\ \alpha_{\mathrm{p}_z}^{\mathrm{A}} \\ \alpha_{\mathrm{p}_z}^{\mathrm{B}} \end{pmatrix} = \mathrm{E} \begin{pmatrix} \alpha_{\mathrm{s}}^{\mathrm{A}} \\ \alpha_{\mathrm{s}}^{\mathrm{B}} \\ \alpha_{\mathrm{p}_x}^{\mathrm{A}} \\ \alpha_{\mathrm{p}_x}^{\mathrm{B}} \\ \alpha_{\mathrm{p}_y}^{\mathrm{A}} \\ \alpha_{\mathrm{p}_y}^{\mathrm{B}} \\ \alpha_{\mathrm{p}_z}^{\mathrm{A}} \\ \alpha_{\mathrm{p}_z}^{\mathrm{B}} \end{pmatrix}$$

Solving this eigenvalue problem gives the *simplest* orbital energy description. Note how we chose to order the orbitals. With this choice, the matrix decomposes into four 2 × 2 eigenvalue problems, which can easily be solved. For example, we have

$$\begin{pmatrix} E_{\mathrm{s}} & \mathrm{s^A s^B} \\ \mathrm{s^A s^B} & E_{\mathrm{s}} \end{pmatrix} \begin{pmatrix} \alpha_{\mathrm{s}}^{\mathrm{A}} \\ \alpha_{\mathrm{s}}^{\mathrm{B}} \end{pmatrix} = \mathrm{E} \begin{pmatrix} \alpha_{\mathrm{s}}^{\mathrm{A}} \\ \alpha_{\mathrm{s}}^{\mathrm{B}} \end{pmatrix}$$

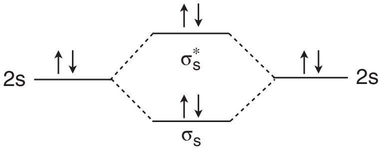


Figure 11.1 Bonding and antibonding states involve atomic s orbitals: σ_s and σ_s^* states, respectively.

Our solution for the eigenvalues cannot be much easier: $E = E_s \pm s^A s^B$. We define the energy $E_s - s^A s^B$ to be σ_s and the energy $E_s + s^A s^B$ to be σ_s^*. In Figure 11.1, we illustrate the splitting of the 2s states.

If we were to think about He_2 instead of C_2, we could immediately see why a He_2 molecule cannot form a stable diatomic molecule. Each orbital can only hold two electrons; one with spin up, the other with spin down as illustrated in Figure 11.1 (in the figure replace 2s by 1s to go from C to He). For He_2, the change in the sum of the eigenvalues vanishes: $\sigma_s^* + \sigma_s - 2E_s = E_s + s^A s^B + E_s - s^A s^B - 2E_s = 0$. Within this simple picture, the electronic energy contribution cancels out. This is not a definitive argument as we made some highly simplified assumptions such as neglecting a number of matrix elements and we took the electronic energy as the sum of one-electron energies, which we know is not rigorous. Still, this argument holds and it explains why He_2 molecules do not form.

We can do the same for the $p_x^A p_x^B$ matrix elements.

$$\begin{pmatrix} E_p & p_x^A p_x^B \\ p_x^A p_x^B & E_p \end{pmatrix} \begin{pmatrix} \alpha_{p_x^A} \\ \alpha_{p_x^B} \end{pmatrix} = \mathrm{E} \begin{pmatrix} \alpha_{p_x^A} \\ \alpha_{p_x^B} \end{pmatrix}$$

Again, we generate a simple solution for the eigenvalues: $E = E_p \pm p_x^A p_x^B$. We define the energy $E_p - p_x^A p_x^B$ to be σ_p and the energy $E_p + p_x^A p_x^B$ to be σ_p^*. The two remaining 2 × 2 blocks yield $E_p \pm p_y^A p_y^B$ and $E_p \pm p_z^A p_z^B$. The energy of these states is degenerate. We label $E_p - p_y^A p_y^B$ by π_p and $E_p + p_y^A p_y^B$ by π_p^*.

In Figure 11.2, we illustrate the energy states for the p-state interactions. A few comments are in order. First, this diagram is *schematic*. We assumed that the π_p level is below the σ_p state. This order is not obvious by anything we have said or written. The level order is important in the sense that if the π_p level is filled, then the dimer cannot have a net magnetic moment. If the σ_p level is lower than the π_p, then this level is filled, and the π_p is only half full. From Hund's rules, we expect the two π_p electrons to have parallel spins, which would produce a magnetic molecule.

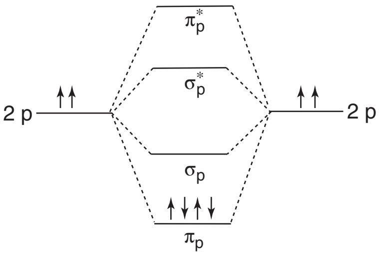


Figure 11.2 Schematic energy levels for σ_p, σ_p^*, π_p, and π_p^* orbitals.

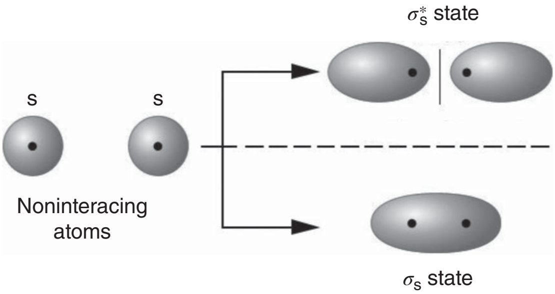


Figure 11.3 Schematic molecular orbital for interacting s states illustrating the origin of the bonding σ_s orbital and the antibonding σ_s^*.

The spatial configuration of the molecular orbitals can be understood in a schematic sense. For C_2 molecule, the s and p orbitals are notably removed in energy. We can discuss their interactions separately.

In Figure 11.3, we consider two s-like orbitals that interact, e.g. consider the 2s states of a carbon atom. There are two orbital configurations from combining s states: One where the states are added together corresponding to the bonding molecular orbital (σ_s) and one where the states are differenced corresponding to the antibonding molecular orbital (σ_s^*). A key difference between the σ_s and σ_s^* orbitals is the presence of a nodal plane in the antibonding orbital.

As one might expect, the p states are more complex because these states have directional axes associated with the orbital. We can think about this as follows. Assume the C_2 molecule is aligned along the x-axis. The p_x states on each C atom are aligned on the axis joining the two atoms. The p_y and p_z orbitals are perpendicular to the x-axis. By symmetry the interactions with p_y, p_z are identical. Again, we do not consider s–p interactions and we neglect differential interactions such as p_x with p_y, or p_y with p_z, i.e. we assume zero differential overlap.

In Figure 11.4, we illustrate the spatial extent of the p_x orbitals. The σ_p orbital is easy to understand as coming from the overlap of p_x-states, which are directed at one another. The π_p orbital is more subtle. The overlap comes from the p_y orbitals (or p_z orbitals) that are perpendicular to the molecular axis. This orbital possesses a node running down the axis. A schematic of this bonding configuration is given in Figure 11.5.

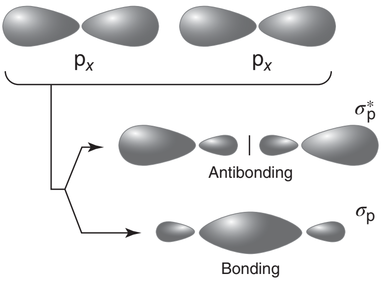


Figure 11.4 Schematic molecular orbital for interacting p_x states illustrating the origin of the bonding σ_p orbital and the and antibonding σ_p^*.

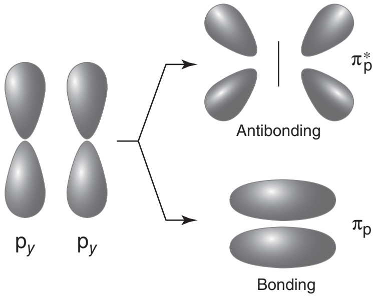


Figure 11.5 Schematic molecular orbital for interacting p_y states illustrating the origin of the bonding π_p orbital and the and antibonding π_p^*.

11.3 Molecular Orbitals: Not Simplified

Computational methods can be used to assess whether our simplified picture is too simple. We take heed of Einstein's admonition: "Make things as simple as possible, but not simpler."

In this section, we illustrate a more realistic (not so simplified) description of the electronic structure of a diatomic molecule. As outlined in Chapter 9, we can apply practical quantum methods to molecular system using pseudopotentials coupled to density functional theory. Specifically, we calculate the spatial and energetic distributions of molecular orbitals.

Our spotlight will again be on the carbon dimer. A short overview of the C_2 molecule (or "dicarbon") can be found in the literature [1]. The molecule is not stable like O_2 or N_2, but it is easily produced in arc experiments, just as C_{60} is. C_2 is also found in comets and is responsible for the blue light we see in flames.

Although only two atoms are involved, dicarbon is a "tough" case to compute as the π_p and σ_p orbitals are nearly equal in energy. As usual, we need to fill each orbital starting from the lowest energy ones. If the π_p orbitals reside below the σ_p orbital, we can fully occupy the π_p orbitals, which can accommodate four electrons. The σ_p orbital is empty. Conversely, if the σ_p state is lower in energy, it will be fully occupied and the π_p states will be half occupied. As we noted before, this filling order will result in "magnetic" carbon according to Hund's rules.

Self-consistent field calculations can be difficult to converge in such situations as "level flipping" occurs. Here is what we mean. Initially, the π_p state might be lower in energy than the σ_p state. We occupy the π_p state with four valence electrons and leave the σ_p empty. In the next iteration, suppose the energy of the π_p state increases and the energy of the σ_p state decreases such that the levels "flip." In this scenario, the π_p state is now partially filled and the σ_p state completely filled. The two states can oscillate in energy as a self-consistent cycle is performed without a stable solution being found.

We need to come up with a workaround to get a solution. Fortunately, an answer exists within density functional theory. Density functional theory does not mandate that each state be integrally occupied. One can fractionally occupy a state. The incorporation of a "fractional" electron occupying a given orbital is not really a flaw of density functional theory. The total density of a molecule is an observable, but not the wave functions, or the individual orbitals.

The occupancy numbers can then be varied until the lowest energy solution is obtained. For example, we could constrain the occupation with "half an electron" in one orbital and half an electron in another, independent of the orbital energies. Using this constrained occupation, we iterate until the system is self-consistent. We then evaluate the total energy. Now suppose we change the constrained occupation. For example, suppose we occupy one orbital by three quarters of an electron and the other by one quarter of an electron. We again iterate to self-consistency with the new occupation and compute the total energy. By exploring the occupancy fractions and evaluating the total energy for every possible occupancy, or at least some subset around the usual integral occupancy prescription, we can find the lowest energy occupation. This is the "correct" solution within density functional theory.

While it seems natural for us to treat the Kohn–Sham orbitals and energy levels in density functional as representing the spatial arrangement and energy of individual electrons, this is not justified by the physics. Nonetheless, the orbitals and energy levels can give us clues as to what to expect for the actual solution.

Consider again the carbon dimer. If the σ_p state is lower in energy than the π_p state, we would expect the molecule to have unpaired spins and from Hund's rules, we might expect the ground state of the molecule to be a spin triplet (or a $^3\Pi$ state). In contrast, if the σ_p state is above the π_p state, then π_p state will be filled and the molecule will not possess a net spin. It will be in a singlet state (or a $^1\Sigma$ state). Theoretically, this is difficult to predict. Experimentally, the ground state of the carbon dimer does not carry a spin and is in the $^1\Sigma$ state with the $^3\Pi$ state only slightly higher in energy [2].

In Figure 11.6, we present contour maps for the molecular orbitals and the total charge density of the carbon dimer. The resulting orbitals bear a remarkable resemblance to the schematic orbitals in Figures 11.3–11.5. The simplified

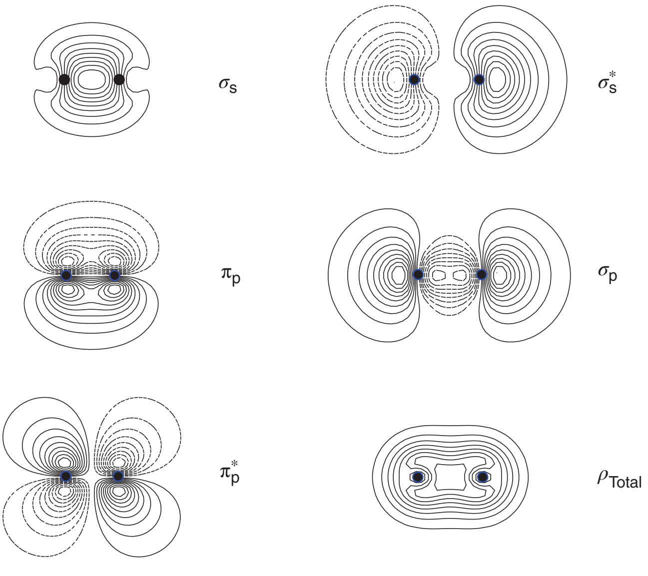


Figure 11.6 Molecular orbitals for the carbon dimer. Negative contours are indicated by dashed lines. The positive/negative phase choice is arbitrary. The positions of the carbon atoms are indicated by black dots.

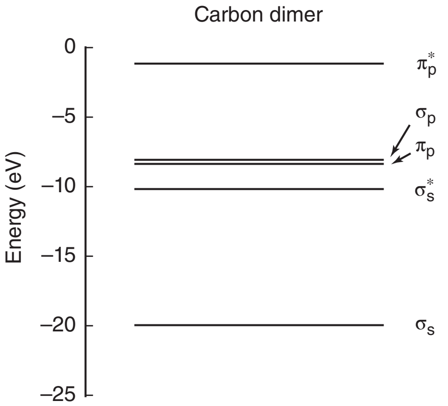


Figure 11.7 Molecular orbital energy levels for the carbon dimer. The π_p and π_p^* states are doubly degenerate. For this configuration, the π_p states are fully occupied and the σ_p state is empty.

picture is at least qualitatively correct compared to more realistic calculations and helps us interpret them.

Our simple picture is less "simple" when the molecular orbital descriptions include d or f states. There can be mixing between the atomic states, and it can be difficult to interpret the contour maps.

Although the energy levels from density functional theory may not be "physical," they can often be associated with real energy levels. For example, if one looks at photoemission spectra using the energy levels as a measure of the binding energy of an electron, the results are often qualitatively correct, if not quantitatively close, to the energies predicted by the Kohn–Sham eigenvalues. In Figure 11.7, the energy levels for the carbon dimer are illustrated. The π_p orbital is doubly occupied, resulting in a closed shell configuration, i.e. all the occupied orbitals are completely filled as expected from experiment.

11.4 Total Energy of a Molecule from the Kohn–Sham Equations

Once the Kohn–Sham energy levels and wave functions are known, one can compute the total electronic energy of the molecule. For the example at hand, it is relatively easy to compute the total energy as a function of the C—C bond length to determine the vibrational frequency of the C—C stretch mode, the equilibrium bond length, and the cohesive energy of the molecule.

We add up the total electronic energy from

$$E(R) = \frac{Z_{\text{ion}}^2 e^2}{R} + \sum_i E_i - \frac{1}{2}\int V_{\text{H}}\,\rho\,\mathrm{d}^3r + \int [E_{\text{xc}} - V_{\text{xc}}]\,\rho\,\mathrm{d}^3r \qquad (11.11)$$

where Z_{ion} is the ionic charge on the C atom (in this case $Z_{\text{ion}} = 4$), R is the distance between the carbon nuclei, the sum is over the occupied states, E_i are the Kohn–Sham energy levels, V_{H} is the Hartree potential, ρ is the charge density, E_{xc} is the exchange–correlation density, and V_{xc} is the exchange–correlation effective potential.

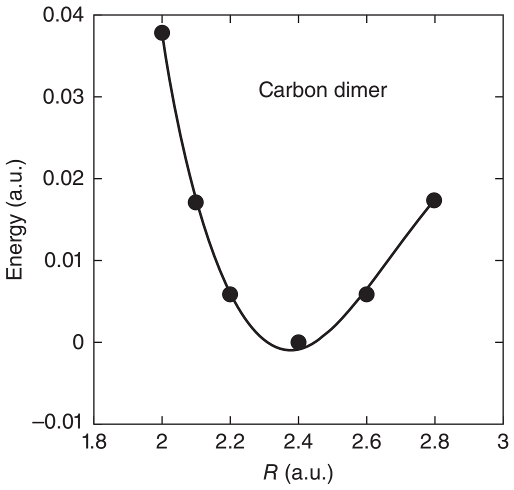


Figure 11.8 Total electronic energy of the carbon dimer per atom as a function of the bond length, *R*. The plot is in atomic units and the energy referenced to the total energy of the molecule.

In Figure 11.8, we plot $E(R)$ vs R for regularly spaced points. Here we consider six points around the lowest energy of the molecule and fit a polynomial to the calculated points. It is often easiest to do this in atomic units (length in Bohr radii, energy in Hartrees – $e = \hbar = m = 1$) and then convert to the desired units at the end of the computation. In atomic units, the fit is given by

$$E(R) = 4.0829 - 4.7034R + 1.7882R^2 - 0.022393R^3 \tag{11.12}$$

The energy is per atom. The order of the polynomial is usually taken to be cubic, which allows for some anharmonicity. To find the equilibrium point R_0, one finds the point where $dE/dR = 0$. This also allows one to find the equilibrium energy, which is referenced to zero in the figure. In pseudopotential calculations, such as this one, the absolute energy of the "pseudoatom" has no meaning. Only relative energies have meaning. The vibrational mode of the molecule can readily be found, assuming that near the equilibrium point, the molecule behaves as a harmonic oscillator. The energy of the oscillator can be written as

$$E(R) = \frac{1}{2}\mu v^2 + \frac{1}{2}kR^2 \tag{11.13}$$

where μ is the reduced mass of the dimer, v is the velocity, and k is the "spring constant" of the molecule. The spring constant can be determined by finding $\mathrm{d}^2E/\mathrm{d}R^2$ at the equilibrium bond length.

In Table 11.1, we summarize the calculated molecular properties compared to experiment. We provide these numbers to illustrate the accuracy of pseudopotentials constructed within density functional theory. Density functional theory is not perfect. Typically, the bond lengths are off by 1–2%, as is the case here. The cohesive energy is the most problematic quantity and bears special mention. The total energy of our carbon pseudoatom is about −145.5 eV, and again, depends on the pseudopotential and the functional. The energy per atom of the carbon dimer is about −150.3 eV. This leaves us with a value of 4.8 eV per atom or 9.6 eV for the cohesive energy of the molecule. The error, if we stopped at this point, would be almost 50%.

Table 11.1 Properties of the carbon dimer; experiment compared to theory.

Property	Experiment	Theory
Bond length (Å)	1.242	1.25
Cohesive energy (eV)	6.31	6.9
Vibrational energy (cm^{-1})	1854.7	1839

However, we have underestimated the energy of the pseudoatom by using a functional based on local density functional that does not account for the spin states of the atom. The carbon atom has a net spin in the ground state. Recall from Hund's rule (in Chapter 10) that the carbon should have a net spin. This additional degree of freedom lowers the energy of the atom by almost 1.3eV. Since we have two atoms, the cohesive energy is reduced to about 7.0 eV. We can improve on this by considering the zero point motion of the molecule. This is not a large correction, but it helps move the theoretical values toward experiment.

Recall that the vibrational modes of a harmonic oscillator are quantized:

$$E_n = (n + \frac{1}{2})\hbar\omega \tag{11.14}$$

where $n = 0, 1, 2, \ldots$. Since $E_0 = \hbar\omega/2$, the energy of the molecule should be raised above the minimum shown in Figure 11.8, even at absolute zero. The vibrational energy of 1839 cm^{-1} is about 0.2 eV (see Appendix A for energy units) and reduces the cohesive energy to roughly 6.9 eV. An error of about 10% for the cohesive energy is actually not that bad. Often, the errors within the local density approximation are notably larger than this error. For this reason, more advanced functionals have been developed such as the generalized gradient approximation, which notably improves the cohesive energy errors of density functional [3].

Example 11.1 ***Water*** For a diatomic molecule, there is only one vibrational mode: stretching the bond length. However, if we have a more complex molecule, there can be numerous degrees of freedom. The number of degrees of freedom can be determined by some simple counting arguments. Suppose we have N atoms in a molecule. If $N = 2$, we would have $3N = 6 = 3 + 2 + 1$ degrees of freedom. Three degrees of freedom can be associated with the center of mass motion. There are two degrees of freedom associated with the rotational modes of the molecule (two perpendicular to a line connecting the atoms) and one degree of freedom for vibrational modes. If $N > 2$ and the molecule is not linear, we can write $3N = 3 + 3 + (3N - 6)$ – again, three degrees of freedom for the center of mass, three degrees for freedom for rotational modes, and $3N - 6$ for vibrational modes.

We can illustrate this by considering a water molecule. Water is a nonlinear molecule. From our counting argument, we should have three vibrational "normal modes." These normal modes can be found if the potential energy surface of the molecule is known. Here we denote the potential energy surface by U. (U is the total electronic energy plus the Coulomb repulsion from the ion cores. Before we called this $E(R)$ for the dimer.) We can generally write the following expansion:

$$U(s_1 \dots , s_n) = U(s_1^0, \dots , s_n^0) + \sum_{j=1}^{n} \sum_{\alpha=1}^{3} \left(\frac{\partial U}{\partial s_{j,\alpha}} \right)_0 \left(s_{j,\alpha} - s_{j,\alpha}^0 \right) + \frac{1}{2} \sum_{j,k=1}^{n} \sum_{\alpha,\beta=1}^{3} \left(\frac{\partial^2 U}{\partial s_{j,\alpha} \partial s_{k,\beta}} \right)_0 \left(s_{j,\alpha} - s_{j,\alpha}^0 \right) \left(s_{k,\beta} - s_{k,\beta}^0 \right) + \cdots \tag{11.15}$$

A word about notation: Here s is a vector: $s = (x, y, z)$, so the jth particle would have $s_j = (x_1, y_1, z_1)$ with the (x, y, z) components labeled by α, e.g. $s_{j,\alpha=1} = x_j$. We choose the equilibrium point to be labeled by s_j^0. At this point the value of U is a minimum and no forces exist on the atoms. In this case, $\left(\frac{\partial U}{\partial s_{j,\alpha}} \right)_0 = 0$ and if we ignore higher order terms we have the equation for a simple harmonic oscillator. Suppose we assume solutions of the form

$$s_j = \frac{1}{\sqrt{M_j}} s_j^0 \exp(i\omega t)$$

where the mass of the jth atom is given by M_j. From Newton's laws, this results in an eigenvalue problem of the form

$$\omega^2 \, s_{j\alpha}^0 = \sum_{j'=1}^{n} \sum_{\alpha'=1}^{3} \frac{1}{\sqrt{M_j M_{j'}}} \left(\frac{\partial^2 U}{\partial s_{j,\alpha} \partial s_{j',\alpha'}} \right)_0 s_{j'\alpha'}^0 \tag{11.16}$$

In our example, water has three atoms so we have to diagonalize a 9 × 9 matrix, called the dynamical matrix. From this diagonalization, one can guess there must be some degeneracies as we know that we only get three normal vibrational modes.

The potential energy surface is the key ingredient from which we can compute the normal modes. We can proceed in several ways.

We could systematically move the atoms in the molecule to map out a potential energy surface and find the derivatives of U. For our water molecule example, we can use standard finite difference methods to extract the needed derivatives. If we want to find $\partial^2 U / \partial x^2$, we can use a formula such as

$$\frac{\partial^2 U(x, y)}{\partial x^2} = \frac{U(x + \Delta x, y) - 2U(x, y) + U(x - \Delta x, y)}{\Delta x^2} \tag{11.17}$$

or with a mixed derivative

$$\frac{\partial^2 U(x, y)}{\partial x \partial y} = \frac{U(x + \Delta x, y + \Delta y) - U(x + \Delta x, y - \Delta y) - U(x - \Delta x, y + \Delta y) + U(x - \Delta x, y - \Delta y)}{\Delta x \Delta y} \tag{11.18}$$

For a water molecule, we can focus on an interatomic potential with only two coordinates, e.g. $U(x, y)$, as the three atoms in water form a plane. We need to

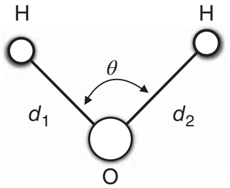


Figure 11.9 Model of a water molecule characterized by two bond lengths, (d_1, d_2) and the H—O—H bond angle, θ.

compute a 6×6 matrix as we have three atoms and find derivatives such as $\partial^2 U/\partial x_1 \partial x_2$ and $\partial^2 U/\partial x_1 \partial y_1$.

One could make a case for extracting these derivatives by taking finite differences of the forces as determined from the Hellmann–Feynman theorem. However, it is often easier to implement the energy differences rather than compute forces. The computation can be simplified by considering symmetry and if the degrees of freedom are not too many, one can plot the potential energy surfaces, fit analytical functions, and take derivatives.

This procedure is perfectly acceptable. Nonetheless, it is easier to examine coordinates tailored for the problem at hand as shown in Figure 11.9. In particular, the H—O—H bond angle, θ, and two H—O bond lengths (d_1, d_2) can be used to define the structure of a water molecule. Using these "collective coordinates" is more physical than using $x_1, y_1, z_1, x_2, y_2, \ldots$.

Still, even this problem requires some work, as it is hard to depict a three variable function, $U = U(d_1, d_2, \theta)$, in two-dimensional plots. We resolve this issue by fixing one bond length, say d_2, at its equilibrium value and plotting U as we move away from the equilibrium values of d_1 and θ. In Figure 11.10, we show the potential energy surface for such a construction.

As one might guess, the energy change for a 2% change in the bond length is larger than a 2% change in the bond angle. Qualitatively, this is a reflection of bond angles being softer, i.e. easier to deform than bond lengths. Also, it takes more energy to compress a bond length by 2% than to stretch it by 2%, which means the H—O bond length should increase with increasing temperature.

A solution of the dynamical matrix eigenvalue problem for the water molecule yields three normal modes. Sometimes it is easy to guess at these modes; the key attribute of the motion in these modes are that the center of mass does not move. The atomic motion of these modes is illustrated in Figure 11.11.

Computed and measured frequencies for these vibrational modes are given in Table 11.2. The level of accuracy is typical for local density approximation calculations. The computed bond length and vibrational modes are off by a few percent from experiment. Since the energy calculations for these modes involve energies for similarly coordinated species, error cancelation is quite good.

The general approach illustrated for the water molecule can be extended in a straightforward manner to more complex systems. However, as the number of atoms increase, so do the number of degrees of freedom. For a system with hundreds of atoms, the vibrational spectra can be quite complex and computationally intensive to compute.

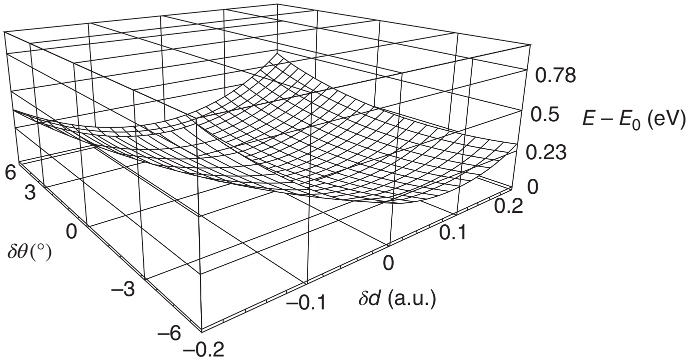


Figure 11.10 Ground-state potential energy surface for water as a function of one bond length and angle, with the second bond length held constant. The calculated equilibrium bond length is 1.81 a.u. and the angle is 104.5°. The corresponding experimental values are 1.81 a.u. and 104.2° [4].

Figure 11.11 Three normal vibrational modes for a water molecule. The motions shown by the arrow should preserve the center of mass of the molecule.

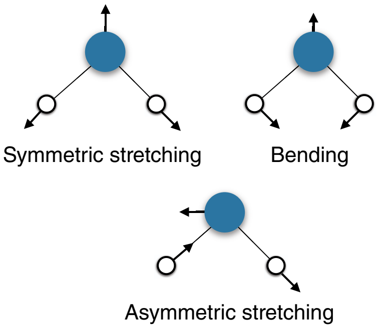


Table 11.2 Properties of the water molecule – experiment compared to theory.

			Vibrational modes		
	Bond length	**Bond angle**	**Symmetric**	**Asymmetric**	**Bending**
Experiment	1.81	104.5	3657.0	3755.7	1594.7
Theory	1.84	104	3667	3811	1606

The bond lengths are in a.u. and the bond angles are in degrees. The vibrational modes are in cm^{-1}. Experiment is from Ref. [4] and theory from Ref. [5].

11.5 Optical Excitations

The “secret” to understanding the electronic and structural properties of molecules is “light.” As in the case of atoms, we know more about molecular properties by analyzing optical spectra than by other means. We focus on *electronic excitations* in this section as the electronic excitations of molecules often fall in the visible spectrum, although it is possible to use photons to excite vibrational or rotational modes.

We can employ theoretical approaches similar to those used for atoms to estimate the excitation energies in molecules. There is one patently obvious difference in molecules when compared to atoms. Molecules possess an atomic structure that can change during an electronic excitation. As a simple example, consider a diatomic molecule. If we remove an electron from a ground-state orbital and place it in an excited-state orbital, the bond is weakened and the bond length of the diatomic molecule will usually expand if the electron remains in the excited state. (Dicarbon is a notable exception to this rule [1].)

The absorption of a photon by a molecule raises a number of fundamental questions. Some of them include: What exactly happens as the molecule absorbs the photon? Does our molecule exist as the superposition of several states when excited, which can only be determined if we attempt to make some sort of measurement? (This is similar to the Schrödinger cat problem.)

We will not address such questions here, save to say the transition can be really fast. The "Franck–Condon" principle captures this temporal disparity. The classical version of this principle states that the rearrangement of the electronic charge is much quicker than the lumbering motion of the nuclei, similar to the Born–Oppenheimer approximation. As such, the nuclear positions do not significantly change while the electron is being absorbed. If we fix the molecular structure, we can determine the electronic energy levels and use these levels to find the excitation energies. Just for the atom, a formalism such as time-dependent density functional theory can be used to find the excited-state energy levels.

11.5.1 Time-dependent Density Functional Theory

In Chapter 10, we used time-dependent perturbation theory to describe the electronic excitations of an atom. We can build on this approach by considering the time evolution of electronic excitations in a molecule. The starting point of such a calculation is a conventional ground-state density functional. The total energy of the molecule is computed, E_0, as well as the Kohn–Sham wave functions and energy levels $(\psi_l(\vec{r}), \varepsilon_l)$.

Within linear response theory, i.e. we proceed as we did for computing the dielectric response of an atom, the optical excitations can be found from the following eigenvalue problem:

$$\mathbf{Q}\mathbf{F}_I = \Omega_I^2 \, \mathbf{F}_I \tag{11.19}$$

where the Ith excitation energy is given by Ω_I. The matrix elements of $\mathbf{Q}$ are given by

$$\mathbf{Q}_{ij,kl} = \delta_{i,k}\delta_{j,l}\omega_{kl}^2 + 2\sqrt{\lambda_{ij}\omega_{ij}} \, \mathbf{K}_{ij,kl}\sqrt{\lambda_{kl}\omega_{kl}} \tag{11.20}$$

where we have dropped some spin indices for simplicity. $\lambda_{ij} = n_j - n_i$ is the difference between the occupation numbers of the i and j Kohn–Sham states. $\omega_{ij} = \varepsilon_i - \varepsilon_j$ is the energy difference between the i and j Kohn–Sham eigenvalues.

(We are using $\hbar = m = e = 1$, i.e. atomic units – see Appendix A.) **K** is a "coupling matrix," whose elements are given by

$$\mathbf{K}_{ij,kl} = 2 \iint \psi_i^*(\mathbf{r})\psi_j(\mathbf{r}) \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \right) \psi_k(\mathbf{r}')\psi_l^*(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{11.21}$$

where E_{xc} is the exchange–correlation energy, which is a functional of the charge density, ρ. Eq. (11.19) is the Casida equation and this approach is called "time-dependent density functional theory."

The expressions for implementing time-dependent density functional theory look "complicated," and they are. However, it could be worse. A number of simplifying assumptions were employed to get to this point. The most notable is that we assume that the effective exchange–correlation energy is local in both space and time. We did this previously in Chapter 10. Also, we consider the perturbing field to be uniform in space; the length scale of any variation of the external field exceeds that of the molecule of interest.

A solution of Eq. (11.19) yields both eigenvalues, Ω_I, and the corresponding eigenvectors, $\mathbf{F_I}$. The eigenvectors can be related to the transition probability and the relative contributions of each Kohn–Sham orbital.

Example 11.2 ***Excitations in the CO Molecule*** Consider a simple diatomic molecule such as CO. If we solve the Kohn–Sham equation for the equilibrium bond length, we can find the first excited level by solving the Casida equation. This is illustrated in Figure 11.12 where we solve for the excitation energy as a function of the bond length. The excited-state curve is found from $E_1(R) = E_0(R) + \Omega_1(R)$ where E_0 is the ground-state energy, and Ω_1 is the lowest excitation energy. $E_1(R)$ is the total energy of the CO molecule as a function of the bond length. The spectroscopic notation for the ground state is $\mathrm{X}^1\Sigma^1$ and for the first singlet excitation energy $\mathrm{A}^1\Pi$. The force on the atom in the excited state can be found from

$$F_1 = -\frac{\partial E_0}{\partial R} - \frac{\Omega_1}{\partial R} \tag{11.22}$$

The value at which F_1 vanishes determines the bond length for the excited state.

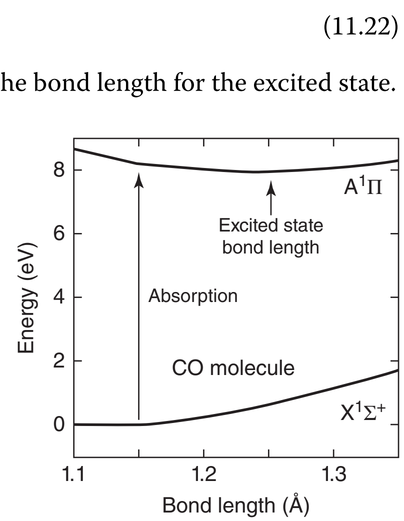


Figure 11.12 Electronic excitation energy for a CO molecule. The ground-state energy for the $\mathrm{X}^1\Sigma^1$ state was computed from a local density approximation. The excited-state energy $\mathrm{A}^1\Pi$ was computed from the time-dependent local density approximation. The absorption energy is computed using the Frank–Condon principle, i.e. the energy is computed at the ground-state equilibrium bond length. The excited-state bond length is also shown.

The bond lengths and excitation energy for the CO molecule are well described by the local density approximation when compared to experiment [6]. The ground-state bond length is calculated to be 1.15 Å, while experiment gives a value of 1.13 Å. The calculated excitation energy and bond length are 7.9 eV and 1.25 Å, compared to experimental values of 8.1 eV and 1.24 Å, respectively. For more complex molecules, the agreement need not be this good. For example, we could get level crossings, i.e. the order of the levels might change with the bond length. If that happens, we cannot compute the force by looking at a single level important interactions.

11.6 Polarizability

Finding the polarizability for atoms is easier than for molecules. The atomic structure of a molecule rules out a one-dimensional description of the electronic structure problem. We need to make sure we have the correct ground-state molecular structure before applying an external electric field. Since we cannot assume an isotropic response, we define a polarizability tensor to account for directional issues:

$$\alpha_{ij} = \frac{\partial \mu_i}{\partial \mathscr{E}_j} = -\frac{\partial^2 E}{\partial \mathscr{E}_i \partial \mathscr{E}_j} \tag{11.23}$$

where μ_i is the dipole along the i-axis ($i = \{x, y, z\}$), $\mathscr{E}_i$ is the field applied along the same axis, and E is the total electronic energy of the system. This expression allows for the real possibility that a field applied in, say, the x-direction can induce a dipole component in the y-direction.

Experimental values are often measured for molecules in the gas phase, so one measures an average over all directions. This average is given by the trace of the polarizability tensor:

$$\langle \alpha \rangle = \frac{1}{3} \mathrm{tr} \langle \alpha_{ij} \rangle = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3} \tag{11.24}$$

A common practice, as for the case of atoms, is to consider a finite field applied to a molecule and find the diagonal elements by finite differencing:

$$\alpha_{ii} = \frac{\mu_i(\mathscr{E}_i) - \mu_i(-\mathscr{E}_i)}{2\mathscr{E}_i} \tag{11.25}$$

where a small applied field, $\mathscr{E}_i$, is applied along and against the i-axis.

Results for a selected set of diatomic molecules and hydrocarbons are given in Tables 11.3 and 11.4, respectively. The calculated values exceed the measured ones as we found earlier for noble gas atoms. In general, the polarizabilities for the larger molecules tend to be in better agreement than those for the smaller ones.

11.7 The Vibrational Stark Effect in Molecules

The traditional Stark effect occurs when energy levels are modified by the presence of an applied electric field. For example, degenerate energy levels may be

Table 11.3 Static electric dipole polarizabilities of selected diatomic molecules calculated using the finite-field methods.

	Polarizability (Å^3)	
Molecule	**Theory**	**Experiment**
H_2	0.91	0.80
O_2	1.64	1.58
NO	1.83	1.70
N_2	1.82	1.74
CO	2.03	1.95
HCl	2.77	2.63
HBr	3.77	3.61
Cl_2	4.73	4.61

Experimental values are taken from several sources: [7]–[10].

Table 11.4 Average static electric dipole polarizabilities of selected hydrocarbons calculated using the finite-field methods

	Polarizability (Å^3)	
Molecule	**Theory**	**Experiment**
Methane (CH_4)	2.65	2.59
Acetylene (C_2H_2)	3.53	3.33
Benzene (C_6H_6)	10.33	10.32
Toluene (C_7H_8)	12.50	11.8

Experimental values are taken from several sources: [10]–[12].

split by the field. Traditionally the Stark effect focuses on electronic energy levels, but vibrational modes can also be altered resulting in an effect in which both electronic levels and vibrational modes are altered by the applied field.

The *vibrational* Stark effect is an important tool for understanding electrostatic environments at the molecular level. The Stark tuning rate describes how much an applied electric field will shift the vibrational frequencies of a molecule. Knowing the tuning rate for a probe molecule then allows one to use spectroscopy to determine the local electric fields the probe is subjected to, which might otherwise not be measurable. In biological systems, this technique is being used to measure the fields generated inside of proteins and nucleic acids by incorporating nitrile probes. In interface science, surface enhanced spectroscopy of self-assembled monolayers on electrodes sheds light on the interfacial potential distribution where local fields may reach 10^7 V cm^{-1}.

In general, we write the tuning rate as

$$hc\Delta\tilde{\nu}_{\text{obs}} = -\Delta\mu_{\text{probe}} \cdot \mathbf{F}_{\text{environment}} \tag{11.26}$$

where $\Delta\tilde{\nu}_{\text{obs}}$ is the change in frequency, $\Delta\mu_{\text{probe}}$ is the tuning rate for a given molecule and vibrational mode, and $\mathbf{F}_{\text{environment}}$ is the electric field. h and c are Planck's constant and the speed of light respectively.

Quantitatively accurate *ab initio* calculations of the Stark tuning rate are challenging. One problem is simply the scale of the effect. A typical probe molecule might have a stretching frequency of roughly 2000 cm^{-1}, which density functional theory can calculate with errors of a few percent. An applied field of 1 MV cm^{-1} would perturb this frequency by less than 1 cm^{-1}. Numerical solutions must be controlled very carefully to detect any frequency shifts.

On a theoretical level, a naive analysis of the vibrational modes might underestimate the size of the tuning rate by an order of magnitude. For example, consider a diatomic molecule such as CO; the vibrational potential, $U(q)$, is perturbed by the product of the dipole moment and the field $\vec{\mu}(q) \cdot \vec{F}$ where q is a position coordinate, in this case the bond length. Assume the vibrational potential is quadratic in coordinate q. In a Taylor expansion of the dipole moment, the constant term $\vec{\mu}(0)$ only changes the total energy. The linear term $\partial\vec{\mu}/\partial q$ shifts the equilibrium bond length, but not the frequency. The quadratic term $\partial^2\vec{\mu}/\partial q^2$ does change the effective spring constant and hence causes a frequency shift, but this explains less than 10% of the tuning rate. When anharmonic effects are considered, the linear term $\partial\vec{\mu}/\partial q$ is responsible for the majority of the change in frequency.

Example 11.3 ***Numerical Computations of the Vibrational Stark Shift*** We illustrate computational methods to compute accurate vibrational Stark shifts for molecules such as CO and benzonitrile (C_6H_5CN). Our focus will be on stretching modes for the C—O or C—N bonds. Our goal is to determine how the stretching modes of these bonds change with an applied field.

We examine three approaches to this problem: (i) a finite field approach, (ii) perturbation theory, and (iii) molecular dynamics simulation.

Finite Field Approach

We have discussed finite fields before to calculate induced dipoles and polarizabilities. The chief difference here is make sure the computations are exceptionally well converged. We need highly accurate computations for the ground-state structure and for the vibrational modes with and without an applied electric field.

To find a particular stretching mode, we begin by constructing the dynamical matrix as in the last chapter:

$$D_{Ii,Jj} = \frac{1}{\sqrt{M_I M_J}} \frac{\partial F_{Ii}}{\partial R_{Jj}} \tag{11.27}$$

where indices I and J run across all atoms, and indices i and j run across the x, y, and z axes. F_{Ii} is the force on the Ith atom in the i direction, and R_{Jj} is the jth Cartesian coordinate of the Jth atom.

Once the dynamical matrix is found, the vibrational modes can be found by diagonalizing the matrix. While the vibrational frequencies can be taken directly from the matrix eigenvalues, these frequencies lack any anharmonic corrections.

Instead, we take the selected eigenmode $\vec{u}$, which is expressed in mass weighted coordinates, and use it to find the atomic displacements $\vec{q}$ for the pertinent stretching mode.

$$q_{Ii}^{n} = \frac{1}{\sqrt{M_I}} u_{Ii}^{n} \tag{11.28}$$

This allows us to find the one-dimensional vibrational potential $U(q)$ as a function of the total electronic energy of the system for finite displacements along the mode $E(\vec{R}_{\text{eq}} + \vec{q})$. As $U(q)$ is a one-dimensional slice of the full potential energy surface $E(\vec{R})$, it picks up the anharmonic components relevant to that mode. To explicitly find the anharmonic vibrational energy levels we construct a 1D Schrödinger equation for this system.

$$-\frac{\mathrm{d}^2\psi_n}{\mathrm{d}q^2} + U(q)\psi_n = \mathscr{E}_n\psi_n \tag{11.29}$$

Bound states of this system are readily found by numerical integration as illustrated in Chapter 7. The frequency of note is given by the fundamental transition from the ground state to the first excited vibrational state $(\mathscr{E}_1 - \mathscr{E}_0)/hc$.

To determine the dependency of this frequency on the electric field, a new vibrational potential $U_{\text{F}}(q)$ is found by repeating the total energy calculations in the presence of a finite field. Again, the vibrational energy levels of this new potential are found by direct integration, and the frequency as a function of field strength is generated. Figure 11.13 illustrates the perturbation of the energy surface and associated vibrational states with an enhanced field to exaggerate the differences. In Figure 11.14, we illustrate how the frequency of the C—O stretch changes as a function of applied field. The tuning rate $\Delta\mu$ is then found by fitting the frequency $\tilde{\nu}(F)$ to the quadratic expression $\tilde{\nu}_0 + \Delta\mu F + \frac{1}{2}\alpha F^2$.

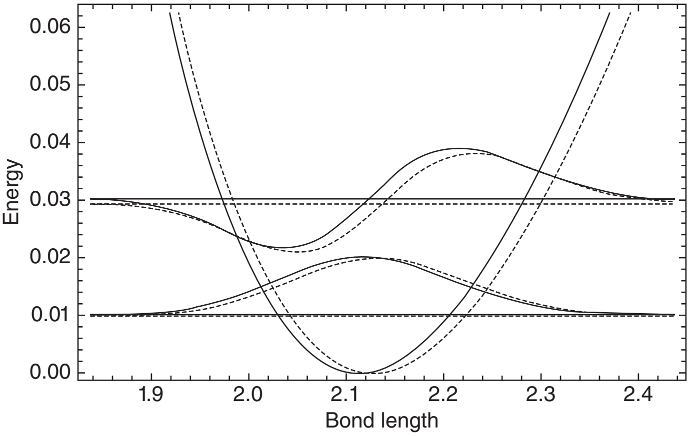


Figure 11.13 Original (solid line) and perturbed potential energy surface (dashed line) under an exaggerated electric field. Shifts in the wave functions and energy levels to the first and second vibrational modes can be observed. This shift as function of the field can be used to compute the tuning rate for the vibrational Stark shift.

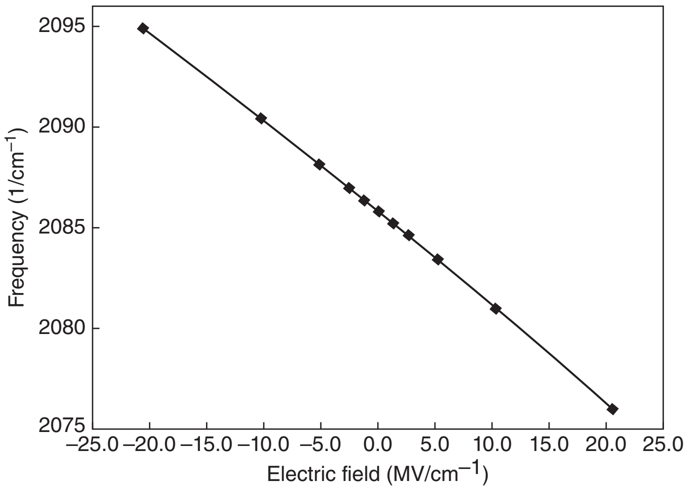


Figure 11.14 Variation of vibrational frequency for the CO molecule as a function of applied field.

Perturbation Theory

Since the tuning rate is defined in the limit of infinitesimal fields, it is natural to also consider a perturbative approach. As before, we construct the dynamical matrix and diagonalize it. We then select the pertinent mode and consider the electronic energy for displacements along this mode $E(\vec{R}_{\text{eq}} + x\vec{q})$ to reduce the problem to a single dimension. In the limit of infinitesimal fields, corrections to this total energy are given by first-order perturbation theory:

$$\Delta E = -\int \vec{r} \cdot \vec{F}\, \rho(\vec{r})\, \mathrm{d}\vec{r} \tag{11.30}$$

Since the applied field is assumed constant this simplifies to

$$\Delta E = -\int \vec{r} \cdot \vec{F}\, \rho(\vec{r})\, \mathrm{d}\vec{r} = -\vec{F} \cdot \int \vec{r}\rho(\vec{r})\, \mathrm{d}\vec{r} = -\vec{F} \cdot \vec{\mu} \tag{11.31}$$

where $\vec{\mu}$ is the dipole moment. The simplest theoretical model for including anharmonic corrections is to consider the harmonic oscillator perturbed by a cubic term.

With this model, "textbook" ladder operators and second-order perturbation theory can be used to find the perturbed energy levels [13, 14]. This approach makes it easy to evaluate "messy" integrals and avoid solving differential equations. The procedure was first developed by Dirac.

Our starting point is the harmonic oscillator Hamiltonian:

$$H(q) = \frac{p^2}{2M} + \frac{M\omega^2}{2}q^2 \tag{11.32}$$

Recall that in quantum mechanics commutation makes a difference. Consider the following commutation with position and momentum. We focus on one dimension:

$$[p, q] = pq - qp \tag{11.33}$$

Classically $[p, q] = 0$, but not quantum mechanically, as p is an operator. The commutation operations have meaning when applied to a wave function:

$$[p, q]\Psi = (pq - qp)\Psi = -i\hbar\frac{\partial(q\Psi)}{\partial q} + i\hbar x\frac{\partial(\Psi)}{\partial q} = -i\hbar\Psi \tag{11.34}$$

We write the expression with the wave function implicit:

$$[p, q] = -i\hbar \tag{11.35}$$

Using this definition, we introduce some useful operators of the form:

$$\begin{aligned} a &= \sqrt{\frac{M\omega}{2\hbar}}\left(q + i\frac{p}{M\omega}\right) \\ a^{\dagger} &= \sqrt{\frac{M\omega}{2\hbar}}\left(q - i\frac{p}{M\omega}\right) \end{aligned} \tag{11.36}$$

This may look strange, but it allows us to write the following relationships involving a and $a^{\dagger}$ based on knowing $[p, q]$:

$$[a, a^{\dagger}] = 1 \tag{11.37}$$

Suppose we have a solution for the harmonic oscillator where the quantum number is given by n. We introduce the bracket notation and can show the following:

$$\begin{aligned} a^{\dagger}|n\rangle &= \sqrt{n+1}|n+1\rangle \\ a|n\rangle &= \sqrt{n}|n-1\rangle \end{aligned} \tag{11.38}$$

We define more operations. Consider the following:

$$\begin{aligned} N &= a^{\dagger}a \\ N|n\rangle &= a^{\dagger}a|n\rangle = a^{\dagger}\sqrt{n}|n-1\rangle = a^{\dagger}\sqrt{n}|n-1\rangle = \sqrt{n}a^{\dagger}|n-1\rangle = n|n\rangle \end{aligned} \tag{11.39}$$

The operator N allows us to write down the solution for the harmonic oscillator as

$$H|n\rangle = \left(N + \frac{1}{2}\right)\hbar\omega|n\rangle = \left(n + \frac{1}{2}\right)\hbar\omega|n\rangle \tag{11.40}$$

This yields the well-known energy levels for the oscillator:

$$E(n) = \left(N + \frac{1}{2}\right)\hbar\omega \tag{11.41}$$

We apply this framework to examine the effect of an anharmonic potential $U(q)$. We modify the Hamiltonian:

$$H_{\text{anh}}(q) = H(q) + Aq^3 \tag{11.42}$$

Writing in terms of ladder operators gives us the following:

$$H_{\text{anh}}(q) = \left(N + \frac{1}{2}\right)\hbar\omega + A\left(\frac{\hbar}{2M\omega}\right)^{\frac{3}{2}}(a^{\dagger} + a)^3 \tag{11.43}$$

The second term is the anharmonic perturbation to the harmonic Hamiltonian. For notation purposes let us define the anharmonic contribution as V. Using perturbation theory we can write

$$E_n^{(1)} = \langle n|V|n\rangle = A\left(\frac{\hbar}{2M\omega}\right)^{\frac{3}{2}}\langle n|(a^{\dagger} + a)^3|n\rangle \tag{11.44}$$

$$E_n^{(2)} = \sum_{m\neq n}\frac{|\langle m|V|n\rangle|^2}{E_n^{(0)} - E_m^{(0)}} = A^2\left(\frac{\hbar}{2M\omega}\right)^3\sum_{m\neq n}\frac{|\langle m|(a^{\dagger} + a)^3|n\rangle|^2}{E_n^{(0)} - E_m^{(0)}} \tag{11.45}$$

The first-order perturbation term, $E_n^{(1)}$, vanishes. We know this because the unperturbed harmonic wave functions are either even or odd wave functions. The integrand is therefore odd (since q^3 is odd) and vanishes when integrated over all space. The second order perturbation term, $E_n^{(2)}$, need not vanish as it can couple two states, one even and one odd, to construct an even integrand.

We are interested in the $n = 0$ and $n = 1$ states as transitions between these two states have the lowest excitation. The second-order term yields the following energies:

$$E_0 + E_0^{(2)} = \frac{\hbar\omega}{2} - \frac{11A^2\hbar^2}{8M^3\omega^4} \tag{11.46}$$

$$E_1 + E_1^{(2)} = \frac{3\hbar\omega}{2} - \frac{71A^2\hbar^2}{8M^3\omega^4} \tag{11.47}$$

As the observable quantity is the energy difference between states, we subtract Eq. (11.46) from Eq. (11.47) to find the fundamental transition energy of

$$\nu = \hbar\omega - 15A^2\hbar^2/2M^3\omega^4 \tag{11.48}$$

This expression allows us to assess the role of a small anharmonic terms on the vibrational modes of a harmonic oscillator. In practice, we do numerical computations to evaluate $U(q)$ and then fit a cubic curve to the results. This fixes the value of A in Eq. (11.42).

We include the effect of the field in terms of the Taylor expansion of the $\vec{F}\cdot\vec{\mu}(q)$ term to generate a new Hamiltonian, H_{F}. This expansion can be determined by numerical computations of the dipole in Eq. (11.31).

$$H_F(q) = \frac{p^2}{2M} + \frac{M\omega^2}{2}q^2 + Aq^3 + F\frac{\partial\mu}{\partial q}q + \frac{1}{2}F\frac{\partial^2\mu}{\partial q^2}q^2 \tag{11.49}$$

With a change of variables, this Hamiltonian can be recast in the same form as Eq. (11.42), with a modified ω_{F} being given by the following:

$$\omega_{\text{F}} = \sqrt{\frac{F}{M}\frac{\partial^2\mu}{\partial q^2} + \sqrt{\left(\omega^2 + \frac{F}{M}\frac{\partial^2\mu}{\partial q^2}\right)^2 - \frac{12A}{M^2}F\frac{\partial\mu}{\partial q}}} \tag{11.50}$$

The frequency in the presence of the field is now given by

$$\nu_{\text{F}} = \hbar\omega_{\text{F}} - 15A^2\hbar^2/2M^3\omega_{\text{F}}^4 \tag{11.51}$$

Taking the linear dependence on F yields the tuning rate:

$$\Delta\mu = -\frac{90A^3\hbar^2 + 3A\hbar M^3\omega^5}{M^5\omega^8}\frac{\partial\mu}{\partial q} + \frac{30A^2\hbar^2 + \hbar M^3\omega^5}{2M^4\omega^6}\frac{\partial^2\mu}{\partial q^2} \tag{11.52}$$

A simple cubic perturbation may not characterize the anharmonicity entirely. Thanks to the machinery of ladder operators, introducing higher order corrections such as a q^4 or q^5 term is possible, although algebraically tedious. The higher order corrections can be found in the literature [15].

Molecular Dynamics

While the perturbative method outlined above provides an additional check on the finite field method, it explicitly depends on the vibrational modes determined by the dynamical matrix. Vibrational frequencies can also be calculated using *molecular dynamics*. This is an attractive option because at no stage must one assume the vibrations are harmonic or that the initial structure is fully relaxed.

Molecular dynamics in its simplest form allows one to follow the trajectories of an atom in a force field. While the forces may be computed from quantum mechanical approaches, the atomic positions are taken as classical coordinates. This works well for the applications here. It does not work so well where quantum effects are present, e.g. at low temperatures.

The trajectory of an atom with an applied force is given by Newton:

$$M_i\frac{d^2\vec{R}_i}{dt^2} = \vec{F}_i \tag{11.53}$$

Our goal is to find the time evolution of the atom's position: $\vec{R}_i(t)$ We have looked at a very similar math problem in a different context – when we solved for the hydrogen atom wave function by direction integration (see Section 7.2).

Consider a simple approach. Again, we split the second-order equation into two first-order equations and solve for the x-component of the atom:

$$\frac{dX_i}{dt} = V_{x,i} \qquad \frac{dV_{x,i}}{dt} = F_{x,i}/M_i \tag{11.54}$$

We can solve for the position using the explicit Euler method (see Eq. (7.23)) by stepping the time by Δt. Let us consider the motion of an atom. We will suppress the atomic index in the following.

$$\begin{aligned}
X(\Delta t) &= X(0) + V_x(0)\ \Delta t \\
V_x(\Delta t) &= V_x(0) + (F_x(0)/M)\Delta t \\
X(2\Delta t) &= X(\Delta t) + V_x(\Delta t)\ \Delta t \\
V_x(2\Delta t) &= V_x(\Delta t) + (F_x(\Delta t)/M)\Delta t \\
&\ \ \cdots \\
X((n+1)\Delta t) &= X(n\Delta t) + V_x(n\Delta t)\ \Delta t \\
V_x((n+1)\Delta t) &= V_x(n\Delta t) + (F_x(n\Delta t)/M)_n\Delta t
\end{aligned} \tag{11.55}$$

Computers are really good at stepping through such equations. We need initial coordinates and velocities for each atom, along with the force for each, i.e. we need $X(0)$, $V_x(0)$ and $F_x(0)$. From the initial conditions, we can find $X(\Delta t)$, $V_x(\Delta t)$. Using the coordinates at Δt, we can find $F_x(\Delta)$ and proceed to find $X(2\Delta t)$, $V_x(2\Delta t)$.

This procedure will work, but it is tedious and can be computationally challenging even for a computer if the time step is small. The validity of the procedure can be checked by examining whether energy is conserved, i.e. the total energy of the atomic motion, including the kinetic and potential energy of the atoms.

This type of molecular dynamics uses the Born–Oppenheimer approximation. For each time step, the electrons respond instantaneously to the positions of the atoms, or better the ion cores.

Operationally, we fix the structure, solve the electronic structure problem, extract the forces, and move the atoms from the equations of motion over a time span, Δt. Take the new structure and repeat (and repeat)!

To start the process, we can assign zero initial velocity to the atoms. The initial stretch would then be a turning point. Or we could imagine some initial velocity set by a specified temperature. This is usually not a crucial issue. Although explicit Euler is not the best way to solve for the equations of motion (other stepping algorithms are more accurate), it is simple and easy to implement.

To excite the relevant modes, we can stretch the bond length of interest and initiate the simulation. For example, suppose we want to follow the C—O bond. We might stretch it along the bond axis by some small fraction of the bond length, say a few percent. We then evolve the system in time using the equations of motion.

A word about time – one atomic unit of time is 0.024 fs. This is the time required for a Bohr electron to travel a distance of one Bohr a_0=0.529 Å, at the velocity of an electron in the lowest state, $v = e^2/\hbar = (e^2/\hbar c)c = c/137.04$, where we noted $e^2/\hbar c = 1/\ 137.04$ is the fine structure constant. The Bohr electron is not moving near relativistic speeds.

A time step for a typical molecular dynamics simulation is about 10 times this. We want the time to be as large as possible to save computer time, but much less than a vibrational mode of the molecule. Otherwise, the simulation would be inaccurate.

In Figure 11.15, we plot the total energy and potential energy of a benzonitrile molecule (C_6H_5CN), where the C≡N bond is stretched. The total energy as a function of time should be a constant. However, the total energy exhibits some very small fluctuations owing to numerical errors. The important issue is that there is no net drift of the energy that would indicate significant errors in the forces and a violation of energy conservation.

Once we initiate motion in the molecule for some given bond length distortion, we record the motion of the atoms involved with and without an applied electric field. In Figure 11.16, the C≡N bond length is plotted as a function of time. Although the difference with and without the field is small and does not show up for several time steps, it is discernible in the plot.

The frequency shift was recovered from the simulation by taking the power spectrum of the C≡N bond length as a function of time and plotting the intensity.

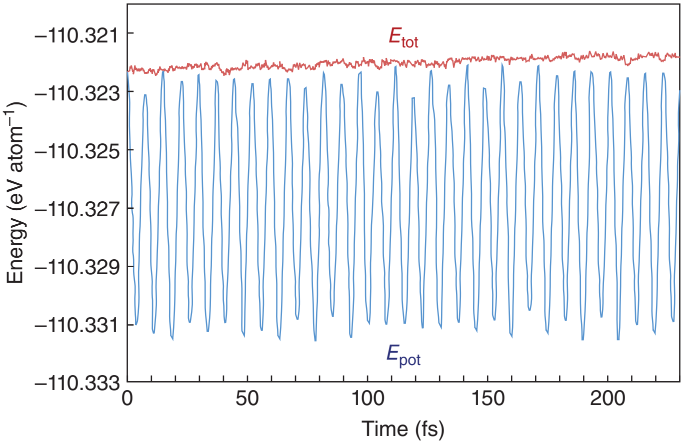


Figure 11.15 Variation of potential and total energy as a function of time for benzonitrile.

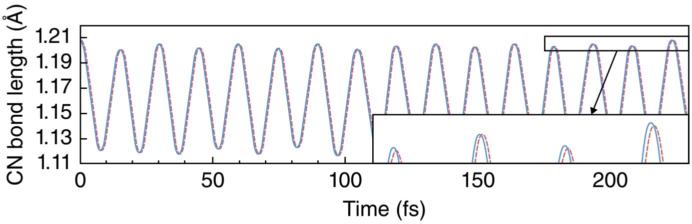


Figure 11.16 Change in the C≡N bond length of benzonitrile as a function of time using a molecular dynamics simulation. Two simulations are shown: one with an applied field (dashed line) and one without (solid line). Differences in the motion, albeit quite small, are clearly altered by the presence of the field.

We use the power spectrum computed earlier to find the dominant frequencies (see Eq. (10.40)). With a minor modification from the previous version of this equation, we can write

$$P(\omega) = \frac{1}{T}\int_0^T \mathrm{d}(t)\exp(-i\omega t)\,\mathrm{d}t \tag{11.56}$$

where T represents the run time of the simulation. The bond length of interest is given by $d(t)$. Suppose we have a "pure" frequency in the bond length frequency: $d = d_0 \exp(i\omega_0 t)$. It is clear that $P(\omega_0) = d_0$. What about $P(\omega \neq \omega_0)$? In the limit of $T \to \infty$, $P(\omega \neq \omega_0) \to 0$, although the integrand is not well defined, we know that it is bounded. Calculating the power allows us to determine what frequencies are present in a straightforward manner. However, the procedure is not flawless. Long run times are required to unambiguously resolve two almost equal frequencies, which is of course our situation as the change in frequencies with and without the field is small. Typically, thousands of time steps are run and a very large electric field is used to exaggerate the differences.

Table 11.5 illustrates computations for the three approaches for the two molecules, CO and benzonitrile, corresponding to the C≡O and C≡N bond

Table 11.5 Comparison of finite field, perturbative, and MD calculations of the tuning rate

Molecule	$\Delta\mu$ Finite field	$\Delta\mu$ Perturbation	$\Delta\mu$ Molecular dynamics	Experiment
CO	0.46	0.44	0.52	0.4–0.67
Benzonitrile	0.33	0.28	0.34	0.6

Units are in $\text{cm}^{-1}/(\text{MV/cm}^{-1})$. Experiment is from [16] and references therein.

stretching modes, respectively. The different approaches agree to within $\sim$ 10%, which is an important conclusion as it confirms the different computational methods. Experimental values can vary notably and often reflect the presence of local fields as when the molecule resides in a liquid.

References

1 Hoffmann, R. (1995). Marginalia: C2 in all its guises. *American Scientist* 83: 309.

2 Huber, K.P. and Herzberg, G. (1979). *Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules*. Van Nostrand Reinhold Company.

3 Becke, A.D. (201). Perspective: fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics* 140: 18A301.

4 Herzberg, G. (1966). *Molecular Spectra and Molecular Structure*, Vol. 3. Van Nostrand.

5 Lopez del Puerto, M., Tiago, M.L., Vasiliev, I., and Chelikowsky, J.R. (2005). Real-space pseudopotential calculations of the ground-state and excited-state properties of the water molecule. *Physical Review A* 72: 052504.

6 Radzig, A.A. and Smirnov, B.M. (1985). *Reference Data on Atoms, Molecules and Ions*. Springer-Verlag.

7 Kolos, W. and Wolniewicz, L. (1967). Polarizability of the hydrogen molecule. *The Journal of Chemical Physics* 46: 1426.

8 Newell, A.C. and Bard, R.C. (1965). Absolute determination of refractive indices of gases at 47.7 Gigahertz. *Journal of Applied Physics* 36: 3751.

9 Maryott, A.A. and Buckley, F. (1953). Table of dielectric constants and electric dipole moments of substances in the gaseous state. U.S. National Bureau of Standards Circular No. 537.

10 Hirschfelder, J.O., Curtis, C.F., and Bird, R.B. (1954). *Molecular Theory of Gases and Liquids*. New York: Wiley.

11 Bose, T.K., Sochanski, J.S., and Cole, R.H. (1972). Dielectric and pressure virial coefficients of imperfect gases. V. Octopole moments of CH_4 and CF_4. *The Journal of Chemical Physics* 57: 3592.

12 Aroney, M.J. and Pratten, S.J. (1984). Electric birefringences and molecular anisotropic polarisabilities of benzene and of nine methylbenzenes and t-butylbenzenes. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* 80: 1201.

13 Griffiths, D.J. (2004). *Introduction to Quantum Mechanics*, 2e. Prentice Hall.

14 Liboff, R.L. (2002). *Introductory Quantum Mechanics*. Addison-Wesley.

15 Garrett, B.F., Azuri, I., Kronik, L., and Chelikowsky, J.R. (2016). Real-space pseudopotential method for computing the vibrational Stark shift. *The Journal of Chemical Physics* 145: 174111.

16 Dalosto, S.D., Vanderkooi, J.M., and Sharp, K.A. (2004). Vibrational stark effects on carbonyl, nitrile, and nitrosyl compounds including heme ligands, CO, CN, and NO, studied with density functional theory. *The Journal of Physical Chemistry B* 108: 6450.

12

Atomic Clusters

> *In physics, the term "clusters" denotes small, multiatom particles. As a rule of thumb, any particle of somewhere between 3 and 3×10^7 atoms is considered a cluster. Two-atom particles are sometimes considered clusters as well.*
>
> Wikipedia

12.1 Defining a Cluster

The Wikipedia definition of a cluster is wrong or at least misleading. Not all multiatom particles are clusters; some are molecules. Consider the following example. Suppose one takes benzene gas molecules (C_6H_6) and condenses the gas into a crystalline state. The crystal structure can be characterized by the arrangement of the molecules, each acting as a C_6H_6 identifiable unit. In contrast, suppose we consider an ensemble of silicon clusters, say Si_{12}, in the gas phase and condense this gas into a crystalline phase. In the ground state of the crystal, we know that the silicon atoms will organize into a diamond crystal. However, within the diamond crystal there is nothing to indicate that a distinct unit of Si_{12}, which existed in the gas, resides within the crystal. One can argue that the different behavior of C_6H_6 and Si_{12} constitutes the fundamental difference between a molecule and a cluster. C_6H_6 is a *molecule* and Si_{12} is a *cluster*. Our definition of a cluster: *A cluster is an assemblage of atoms that is stable only in isolation*. This definition properly distinguishes between a molecule and a cluster.

Experiments on clusters are difficult as one might expect. Measurements on these "fragile objects" must be done when they are in isolation. Two techniques of accomplishing isolation are often employed. One is to embed the cluster in an inert medium, e.g. a condensed inert gas. Since the medium is inert, it fixes the cluster for study, but does not significantly interact with it. Another is to employ "beams" of clusters in which the clusters are kept apart.

Both these experimental techniques possess drawbacks. The inert media technique may not be truly "inert" and might affect the properties of the cluster even in the absence of any chemical bonds being formed. For beam techniques to be effective, the density of clusters must be sufficiently low to prevent collisions or aggregation. Yet, the density of the clusters must be sufficient to allow experimental detection.

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

12.2 The Structure of a Cluster

The difference between a cluster and a molecule becomes quite apparent when one attempts to quantify the structure of each. In our example with silicon clusters and benzene molecules, we could condense benzene into a crystal and use X-rays to determine the crystal structure. Most molecular structures are determined this way. If one can condense the molecule of interest into a crystalline state, one is in business. This is not the case for a cluster. In our example of Si_{12} clusters, we could condense the clusters into a crystal, but any given cluster would lose its identity once condensed.

The lack of an experimental means to measure the structure of a cluster is a serious matter. And, computational techniques are of limited help. How might one determine the structure of a simple cluster? Suppose we make it relatively easy. We consider a very small cluster, say just a few atoms in size.

We could enumerate all topologically distinct clusters. Given a reliable theoretical means for determining the energy of a given geometry, we could compute the structural energy for each and every cluster geometry and find the one with the lowest energy. The key wording is "a very small cluster." A cluster with, say, three or four atoms is one thing. It is not hard to enumerate the distinct geometries. The structure of a five-atom cluster starts to become an issue when one considers the enumeration of all geometries. Enumerating all possible structures of a 10-atom cluster starts to be a formidable problem. Enumerating all possible structures of a 100-atom cluster is beyond being hopeless.

The structural problem centers on the exponential-like growth of possible structures. Some quantities that scale poorly remain "computable." The "pre-factor" is small and the "time to solution" is still a reasonable number. That is not the case here. One might be able to compute the structural energy for a 100-atom cluster on a laptop, but the number of topologically distinct clusters increases astronomically once one exceeds a dozen or so atoms. Even if we did not do quantum calculations for the clusters, the current computational load to find the lowest energy structure is beyond our means.

Enumerating all structures without regard to energetics does not make sense on another level. Some potential structures are sufficiently unfavorable in terms of the total energy that they can be tossed out from the beginning. But, typically a vast number of structures exist that possess nearly degenerate energies. The means to pick *the* lowest energy one is problematic at best.

In short, with the notable exception of trivially small clusters, no method can guarantee that the structure found is in the ground state. Our best hope is to obtain a set of "reasonable structures."

At this point, one could ask: Does it really matter whether we have the *definitive* ground state or not? Probably not. For many situations, clusters do not exist in a state where we have a definitive, isolated, distinct, ground-state structure. Rather, we might consider some set of low energy clusters. For example, suppose we knew most of the structures within ~0.01 eV/atom of the definitive ground-state properties. For many purposes, this is good enough to provide a decent representation of characteristic structures, i.e. a decent ensemble.

In practice, our goal of searching for reasonable structures is to find a subset of low-energy clusters. In carrying out a search for such structures, we want to avoid being trapped in local minima, to proceed in an unbiased manner and to carry out the search in a computationally efficient manner, where one does not spend a great deal of time looking at topologically similar structures.

12.2.1 Using Simulated Annealing for Structural Properties

One of the most annoying issues in a structural search (besides dealing with a large number of topologically distinct clusters) is the possibility of getting localized in an energetically unfavorable, but locally stable, structure. In this situation, one relaxes a potential structure wherein no residual forces exist, but the structure is not in the true ground state. Most search methods center on avoiding this situation.

One attractive method to determine the structure of moderate-sized clusters is *simulated annealing*. This method is analogous to an annealing process that occurs in nature. Consider a solid that is formed by cooling the melt. At temperatures well above the melting point, atoms in the melt will sample a large number of configurations. Bonds are continually being broken and reformed. As the liquid is cooled, unfavorable configurations become less likely and the atoms start to assume a structure closer to the solid state. If one cools very slowly, energetically favorable micro-structures are common and the atoms eventually organize into a crystalline state.

The key issue in simulated annealing is the cooling process. If one does not cool slowly, one might "freeze in" an unfavorable structure. For example, a quickly cooled liquid may not crystallize properly – it may form an amorphous solid with no long-range order, or it may form a crystalline structure with many defects. Physically, this can happen when cooling a real liquid.

The goal of a simulated anneal for cluster formation is similar to that of cooling a liquid. We begin by considering a "random" assemblage of atoms. This initial geometry is maintained at a high temperature by a heat bath and slowly cooled.

At first, energetically unfavorable structures can be accessed as the hot atoms move stochastically, e.g. they execute Brownian motion with random collisions. Only when the temperature of the heat bath is cooled will the clusters be constrained to more favorable structures by forming chemical bonds.

If one cools too quickly, a structure may be overlooked; if one cools too slowly, the same structures can be probed over and over again with no new information obtained. In this sense, simulated annealing is often more of an art than a science. Often one attempts to adjust the temperature so that a number of energetically unfavorable structures are kept in the mix as long as it is computationally possible to do so. If the anneal is done slowly enough, one hopes that the procedure will quench out reasonable candidates for the ground-state structures, but there is no absolute guarantee.

Langevin molecular dynamics is well suited for simulated annealing methods [1], [2]. The French physicist Paul Langevin developed a method that uses stochastic differential equations to model a particle moving in a medium characterized by a temperature and a viscosity. In Langevin dynamics, the atomic positions, $\mathbf{R}_j$, evolve according to

$$M_j\,\ddot{\mathbf{R}}_j = \mathbf{F}(\{\mathbf{R}_j\}) - \gamma M_j\,\dot{\mathbf{R}}_j + \mathbf{G}_j \tag{12.1}$$

where $\mathbf{F}(\{\mathbf{R}_j\})$ is the interatomic force on the jth particle, and $\{\mathrm{M}_j\}$ are the atomic masses. The last two terms on the right-hand side of Eq. (12.1) are dissipation and fluctuation forces, respectively. The dissipative forces are defined by the viscosity coefficient, γ. The fluctuation forces are defined from a random Gaussian distribution, $\{\mathbf{G}_i\}$, with a white noise spectrum:

$$\langle G_i^\alpha(t)\rangle = 0 \quad \text{and} \quad \langle G_i^\alpha(t)G_j^\alpha(t')\rangle = 2\gamma\; M_i\; k_\mathrm{B}\; T\; \delta_{ij}\; \delta(t-t') \tag{12.2}$$

where t denotes time, the angular brackets denote ensemble or time averages, α stands for the Cartesian component, T is the temperature, and k_B is the Boltzmann constant. The coefficient of T on the right-hand side of Eq. 12.2 insures that the fluctuation-dissipation theorem is obeyed, i.e. the work done on the system is dissipated by the viscous medium. The interatomic forces can be obtained from interatomic potentials or from quantum calculations. Quantum-derived forces are always better as they can accurately replicate charge transfer and hybridization changes as bonds form and break in the anneal process.

Choosing an initial atomic configuration for the annealing simulation takes some care. If the atoms are too far apart, they will exhibit Brownian motion, which is appropriate for Langevin dynamics with the interatomic forces zeroed out. In this case, the atoms will not form a stable cluster. As the simulation proceeds, since there is nothing to bind them together, they will simply execute a random walk and disperse. Conversely, if the atoms are too close together, they can immediately form bonds and potentially lock in an unfavorable configuration from the beginning, i.e. they may form a metastable cluster from which the ground state may be kinetically inaccessible even at the initial hot temperature. This brings up another issue. What is a “hot temperature”? In an engineering sense, we want the initial temperature of the heat bath to be sufficiently hot to break and reform bonds. Otherwise, the bonds will form and never break again.

A related and fundamental question centers on the meaning of temperature. Defining the temperature of a cluster is clearly not a straightforward matter. Temperature is usually defined as an average over some macroscopic ensemble and not by summing over a few dozen atoms.

We can circumvent this issue by placing the cluster in a heat bath as defined by Langevin dynamics. There are other techniques for controlling temperatures, e.g. by rescaling the velocities of the atoms in the cluster to agree with some prescribed value. Specifically, the kinetic energy of a cluster can be used as a measure of the temperature. From the equipartition theorem, each degree of freedom can be assigned a value of $k_\mathrm{B}T/2$ where k_B is the Boltzmann constant and T is the temperature.

We make the usual assumption that the atoms in the simulation can be treated as classical objections. This assumption makes sense as long as we do not enter regimes where quantum mechanics starts to play a role such as might be the case for very low temperatures.

The kinetic energy, K, per atom for a cluster would be $3k_\mathrm{B}T/2$. There are several mechanisms for creating an ensemble with a given temperature that reflect the equipartition theorem. A common procedure is to use Langevin dynamics, which controls the temperature by the heat bath. Another common procedure is to assign random velocities to the atoms such that the root mean square velocity is given by $v_\mathrm{rms} = \sqrt{3K_\mathrm{B}T/M}$ where for case at hand, M, is the mass of the atom. As long as energy is conversed the cluster will maintain this temperature.

Often, the initial cluster geometry is formed by a random placement of the atoms with some constraints. The constraints reflect our desire to have the atoms interact, but not so strongly that they form "unbreakable" bonds in the simulation. For example, we might require for a given atom that at least one other atom must reside within, say, 1.05–1.3 times the bond length, where the bond length is defined from experiment or from atomic radii. This constraint means that the atoms are initially weakly interacting and should not drift apart. Another trick is to place the atoms in a spherical domain such that as the atoms move toward the domain boundary a fictitious force appears that drives them toward the center of the domain. This is a reasonable approach as one never loses an atom by diffusion.

There are different ways to integrate an equation of motion. Suppose we consider a Taylor series for the position of an atom, which for simplicity we write in one dimension:

$$x(t+\Delta t) = x(t) + \frac{\mathrm{d}x}{\mathrm{d}t}\Delta t + \frac{\mathrm{d}^2x}{\mathrm{d}t^2}\Delta t^2 \tag{12.3}$$

where the velocity of the particle is given by $v = \frac{\mathrm{d}x}{\mathrm{d}t}$ and the acceleration by $a = \mathrm{d}^2x\mathrm{d}t^2$, with the derivative evaluated at the time, t. If the force on the atom, F, is a simple function of position, then we can use Newton's law (F=Ma) to write

$$x(t+\Delta t) = x(t) + v(t)\Delta t + a(t)\Delta t^2 \tag{12.4}$$

We need to know how to advance the velocity. A common approach is to write

$$v(t+\Delta t) = v(t) + \frac{(a(t) + a(t+\Delta t)}{2}\Delta t \tag{12.5}$$

Note how the acceleration is computed as an average from the initial time and the advanced time.

How does this work in practice? Suppose we know the initial positions and velocities of all the atoms of interest. (Since we are dealing with a second-order equation we need to know both.) The initial positions are usually known or assumed. The velocities can be arbitrarily assigned at random, but constrained so that the kinetic energy is consistent with the desired temperature.

One can step the position of the atom over an interval Δt. The position of the given atom i at x_i at time Δt can be taken as

$$x_i(\Delta t) = x_i(0) + v_i(0)\Delta t + a_i(0)\Delta t^2 \tag{12.6}$$

where $x_i(0)$ is the initial position and $v_i(0)$ is the initial velocity. If we know the initial positions of all the other atoms, $x_j(0)$ $(j \neq i)$, then we know the force on the atom of interest. If we have a classical force field, this is easy to compute. If we have a quantum system, we can use the Hellmann–Feynman theorem. This is not so easy, but given some computing power it is doable. This gives us the acceleration from $a_i(0) = F_i(0)/M_i$. We now know $x_i(\Delta t)$ for all the atoms. To find $x_i(2\Delta)$, we have an issue. We need to know $v_i(\Delta t)$, which we can find from

$$v_i(\Delta t) = v_i(0) + \frac{(a_i(0) + a_i(\Delta t)}{2}\Delta t \tag{12.7}$$

How do we know $a_i(\Delta t)$? Well, if we have the positions of the atoms that were moved to $x_i(\Delta t)$ for all i, we can find the force $F_i(\Delta t)$ and the acceleration from $a_i(\Delta t) = F_i(\Delta t)/M_i$. We keep repeating this process. The "secret" is in choosing Δt such that the positions and forces are accurately computed just as we did in simulating the vibrational Stark shift. Again, one checks this by making sure that energy is conserved as the atoms move.

Langevin dynamics is not very easy to implement. A problem arises because the forces are not determined by the positions of the atoms alone. They depend on the velocity of the atom owing to the dissipative force. We can illustrate this algorithm by considering a particular atom, i, at $x_i(t)$. The position of the atom at time $t + \Delta t$ is given by

$$x_i(t + \Delta t) = x_i(t) + v_i(t)\Delta t + \frac{4a_i(t) - a_i(t - \Delta t)}{6}\Delta t^2 \tag{12.8}$$

where v_i is the velocity and a_i is the acceleration for the ith atom. a_i is given by

$$a_i(t + \Delta t) = \frac{1}{M_i}\{F_i(t + \Delta t) - \gamma M_i u_i(t + \Delta t) + G_i(t + \Delta t)\} \tag{12.9}$$

where $u_i(t + \Delta t)$ is an "intermediate velocity" given by

$$u_i(t + \Delta t) = v_i(t) + \frac{3a_i(t) - a_i(t - \Delta t)}{2}\Delta t \tag{12.10}$$

with the actual velocity given by

$$v_i(t + \Delta t) = \left\{ x_i(t + \Delta t) - x(t) + \frac{2a_i(t + \Delta t) + a_i(t)}{6}\Delta t^2 \right\} /\Delta t \tag{12.11}$$

The interatomic forces, F_i, are straightforward given the atomic positions. Since the net force depends on the velocity, it is a bit tricky to extract the velocity. For this reason, we introduced an "intermediate" velocity U. Most standard algorithms for integrating the equation of motion do not admit the possibility of having dissipative forces. The random forces, G_i, also require some special attention. The force $G_i(t + \Delta t)$ is not correlated with the force $G_i(t)$. The fluctuating forces are taken as Gaussian random variables such that the probability, $P(G)$, of the atom experiencing a force G is given by

$$P(G) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{G^2}{2\sigma^2}\right) \tag{12.12}$$

where the width of the Gaussian is given by

$$\sigma = \sqrt{2\gamma M k_{\mathrm{B}} T/\Delta t} \tag{12.13}$$

There is yet another annoying aspect of Langevin dynamics. How do we start the process? The time step at $t + \Delta t$ requires that we know what went on at t

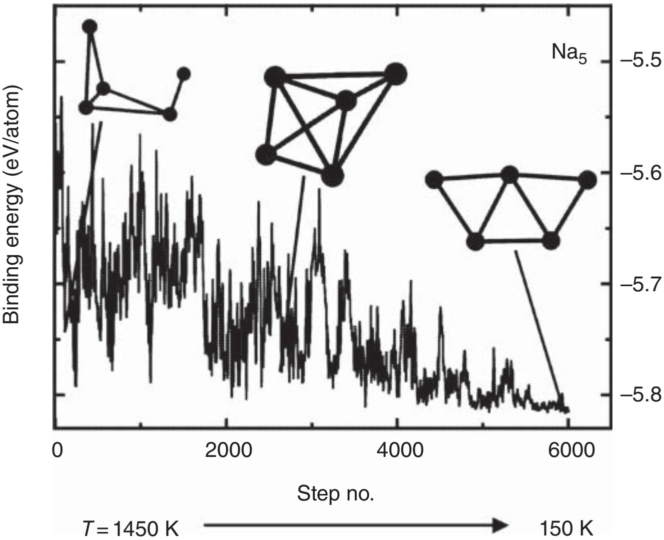


Figure 12.1 Binding energy of Na_5 cluster during a Langevin simulation.

and $t - \Delta t$, at least for the acceleration term. To start the process at $t = \Delta t$, we need to know what goes on at $t = 0$ and $t = -\Delta t$. There are workarounds for this, such as starting the process without the dissipative and random forces, e.g. as we could use Newtonian dynamics and a standard velocity algorithm (known as a velocity Verlet algorithm [3]) for a few time steps and then turn on the heat bath. The key is to run the simulation for a sufficient time such that the initial conditions do not matter. One also has to experiment with the time step such that the system replicates Brownian motion at some temperature in the absence of any interatomic forces.

In Figure 12.1, we illustrate the simulated anneal schedule for a simple cluster: a five-atom sodium cluster. Some of the Na atoms in the early stage of the anneal are one- or twofold coordinated. In the final structure, no Na atom is less than threefold coordinated.

In Figures 12.2 and 12.3, we present the lowest energy structures for Na_n for $9 \le n \le 20$ as determined within simulated annealing. For these modest-sized clusters, we have a decent chance of getting the geometries correct or close to correct, say within ~10 meV/atom. Many studies of Na clusters have shown similar results. Na is a difficult case in that the Na—Na bonds are not highly directional and little energy is lost in changing the bond angle. This is in contrast to covalent systems where sp^3-bonds assume a distinct tetrahedral orientation [4].

12.2.2 Genetic Algorithms

The simulated annealing approach to structure is a very physical approach and works all right for small or medium-sized clusters. However, this approach requires some insight, or many trial and error runs, for larger clusters. The method often wastes time by revisiting the same structures over and over.

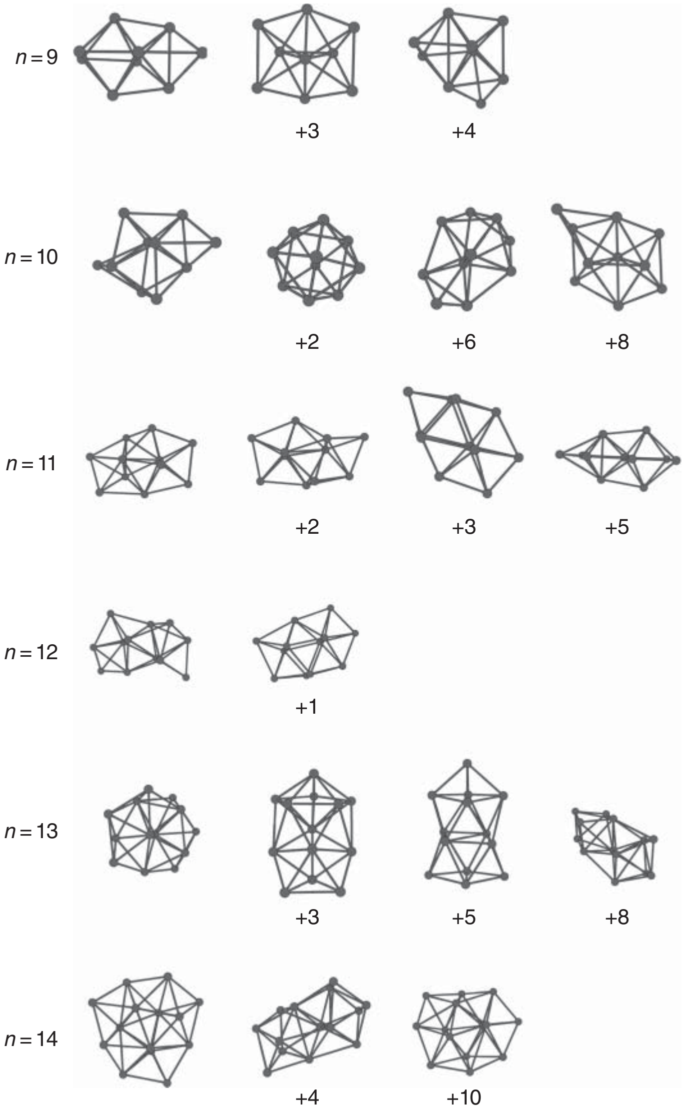


Figure 12.2 Ground-state geometries and some low-energy isomers of Na_n clusters, $n \leq 14$. The energy of the lower lying clusters is given in millielectron volts per atom relative to the likely ground-state geometry.

Genetic algorithms can overcome some of these deficiencies [5]. This algorithm is also based on a physical process related to evolution, i.e. the “survival of the fittest cluster” or better the “survival of the most energetically favorable cluster.” Actually, that is still not quite right. It could be that the most energetically favorable cluster is hard to find and the algorithm might miss it.

Here is how the algorithm works. Consider a couple of candidate clusters chosen by physical intuition, by a simulated anneal run, or perhaps by simply taking a bulk fragment. Imagine that we cut one of these clusters in two, e.g.

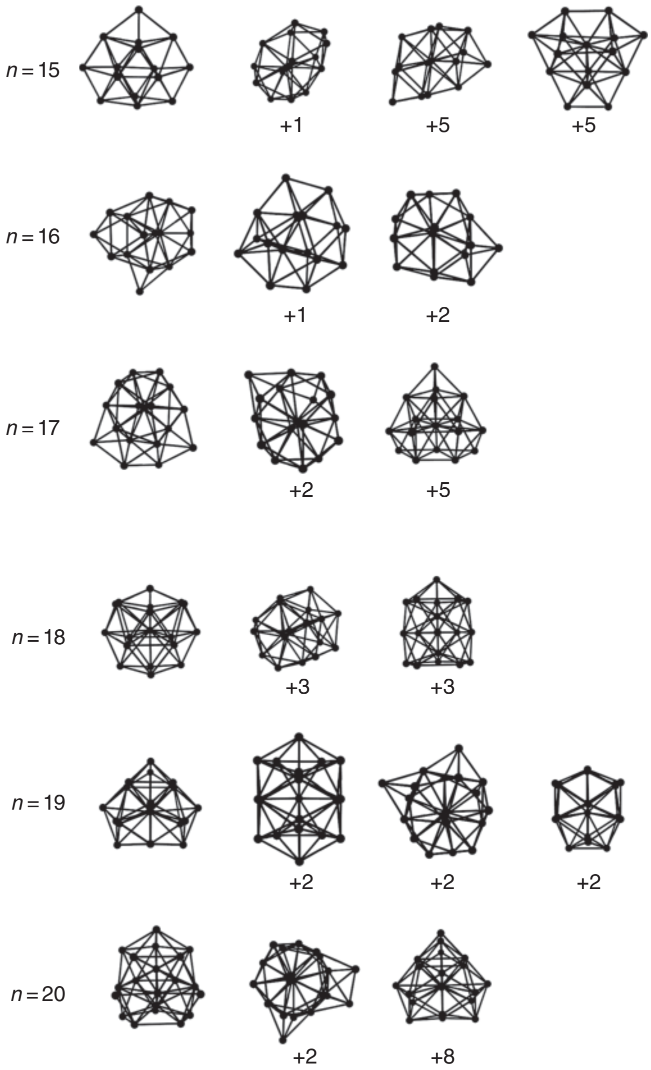


Figure 12.3 Ground-state geometries and some low-energy isomers of Na_n clusters, $n \leq 20$. The energy of the lower lying clusters is given in millielectron volts per atom relative to the likely ground-state geometry.

we might pass a plane through the center of mass of the cluster and create two "half-clusters" with equal number of atoms. We do this again for a different cluster. Taking two "half-clusters," we form new candidate clusters, or in a genetic sense we could call these new candidates: children clusters.

One can see some pesky operational issues with this method. For example, how do we join the half-clusters together? We could orient the half-clusters so that the new arrangement has some reasonable bond lengths (or possibly not) with the hopes of sampling some new geometry regardless of whether the energy is reasonable. Suppose we did this operation in some sensible manner, e.g. we might do some annealing, to create a number of children clusters. We then assess the best ones. "Best" here means the most energetically favorable ones. "Best" in a biological system might mean the one most adapted to reproduce successfully. The analogy to biology is less than perfect.

To be more specific, let us assume we started with four candidate clusters and generated two children clusters from each set of parent clusters. It is not as though our clusters have gender. We could pair a cluster with itself. So each cluster could be paired with itself and three other clusters. Pairing one cluster with another cluster number can result in different geometries as the cutting plane is random; i.e. it need not generate the same half-clusters. We could generate 32 children clusters in this way. From this generation, we could pick the four best and generate 32 more children. We keep doing this generation after generation. We could even choose a less favorable offspring every once in a while to introduce some variety in the process, similar to introducing a mutation in the biological sense.

Some of this is an "art" as there are no strict rules or "guidelines." Many decisions are typically handled by "trial and error." For example, how many children do we need to create from a set of parent clusters? We could take only a few children and do many generations, or take many children and do fewer generations. Also, like simulated annealing there is no well-defined "stopping point." What one typically does is to keep generating offspring and select the ones with the most favorable trait, i.e. the most favorable energy, until one runs out of patience, or computer time, or the structure is so highly symmetric it just looks right! An example follows.

Let us consider 60 carbon atoms and do a genetic algorithm. The generation of possible structures is shown in Figure 12.4. Here 5000 generations were considered starting from a "random" cluster of 60 carbon atoms. After this evolution, a special cluster appeared: a so-called fullerene or "buckyball." The buckyball is special from the viewpoint of its symmetry. Intuitively, it seems doubtful that another more symmetric cluster would exist and support the graphite-like features of the buckyball.

However, one cannot be sure. Suppose we had declared it quits after 4000 generations? We might have missed the special structure altogether. However, given that the buckyball structure exists with defects in earlier generations, one might cut to the chase and remove the defects by hand. One can always look at the total energy for any structure, be it from genetic algorithm or by chance. If the structure is lower in energy, then it is a clear "winner."

An historical note: The discovery of the buckyball, as a new form of carbon, was first inferred by mass spectroscopy experiments in 1985. The structure was later confirmed by crystallization of C_{60} molecules and recognized by a Nobel Prize in Chemistry in 1996. The buckyball is not a cluster that uses our definition. The fact that it can be crystallized into a buckyball solid speaks for its molecular nature. See Appendix B for more on the buckyball.

12.2.3 Other Methods for Determining Structural Properties

"Basin hopping," "big bang," and "particle swarm" methods are also used for doing structural searches.

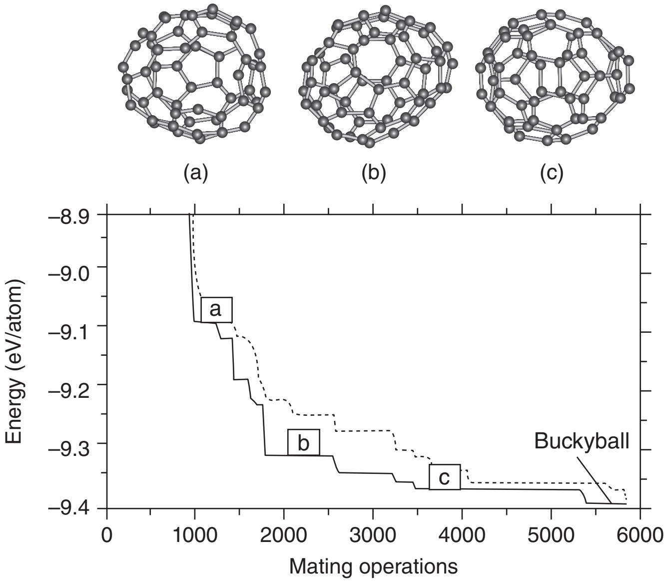


Figure 12.4 Generation of the C_{60} molecule, starting from random coordinates, using the genetic algorithm described in the text with four initial candidates and no mutations. The energy per atom is plotted for the lowest energy (solid line) and highest energy (dashed line) candidate structure as a function of the number of genetic mating operations that have been applied. Several of the intermediate structures that contain defects are illustrated at the top: (a) contains one 12-membered ring and two 7-membered rings, (b) contains a 7-membered ring, and (c) contains the correct distribution of pentagons and hexagons, but two pentagons are adjacent. The ideal icosahedral buckyball structure is achieved shortly after 5000 genetic operations. Source: Figure from Deaven and Ho [5].

Basin hopping centers on running a number of different starting points to avoid being caught in a local minimum [6]. In this scenario, one considers a number of different structures and relaxes each one to a lowest energy structure. If a sufficient number of structures are relaxed one hopes to find the ground state minimum. The probability of finding the ground state structure is proportional to the size of the catchment basin of the "global minimum" compared to the size of other local minimum structures. For some systems, the effective size of the global minimum is sufficient so that only a few random structures might be required, but without some "insider" information one cannot know this.

Another method of note rests on considering a random and highly compressed initial state of atoms [7]. The energy of such a state is very unfavorable relative to the bond energy per atom in the ground state. From such a starting point the atoms would quickly expand, hence the name "big bang." In this method, the atoms are relaxed from this starting point so as not to allow them to expand into empty space and be lost forever. The advantage of this method is that the atoms are highly interacting so that all structures are in principle accessible. As usual, a key to the success of this method is to consider a very large number of starting points so that a large phase space is considered.

The particle swarm method is another intriguing approach [8], [9]. If we construct a random set of clusters, there is no guarantee that the set includes a representative distribution of possible starting points. For example, there is nothing to insure that a number of randomly chosen clusters are of diverse symmetry. This is wasteful. We will not learn new information in relaxing similar structures. We want some "diversity." The swarm method considers "families" or "swarms" of similar structures. Each swarm is then optimized independently of the other.

No matter what the method is, it always pays to explore structural energies as quickly as possible. Some workers expedite the search by considering less accurate methods than "standard" density functional methods for candidate structures. For example, suppose one knows the bond length of a particular atomic species from bulk crystals. One could argue that any structure with a bond length shorter than 5% or longer than 10% is not reasonable. One promptly dismisses candidate structures with such anomalous bonds.

A more quantitative procedure is to use a tight binding model wherein the structural energies are "semi-quantitative." Using this approach, one might collect candidate structures within some prescribed energy. For example, suppose that the lowest observed cluster has a binding energy of 5 eV/atom. One might consider all the candidate structures with an energy of, say, 4.5–5 eV/atom. With this inventory, one could then apply a more rigorous method based on density functional theory. The hope is that the ground state structure, albeit with an incorrect total energy, is still captured by the approximate method.

Once the structure of a cluster is known, or reasonably well approximated, electronic and optical properties can be determined.

12.3 Electronic Properties of a Cluster

We can determine the electronic properties of clusters using the same tools we outlined for atoms and molecules. Given a structure, we solve for the spatial and energetic distributions of electrons in a cluster. This solution is used to determine the response of a cluster to electronic or magnetic fields.

12.3.1 The Electronic Polarizability of Clusters

One of the few cluster properties that can be measured with relative ease and accuracy is the static polarizability. α_{ij} is defined as before in Eq. 11.23:

$$\alpha_{ij} = \frac{\partial \mu_i}{\partial \mathcal{E}_j} = -\frac{\partial^2 E}{\partial \mathcal{E}_j \partial \mathcal{E}_j}; \quad i,j = \{x,y,z\} \tag{12.14}$$

A convenient approach for computing the polarization of *confined* systems, such as clusters, is to include the external electric field directly in the Hamiltonian. We did this before for molecules in Section 11.7.1 and briefly review the procedure for clusters. The external ionic potential $V_{\text{ion}}(\vec{r})$ experienced by the electrons is modified to have an additional term given by $-e\vec{\mathcal{E}} \cdot \vec{r}$, where $\vec{\mathcal{E}}$ is the electric field vector.

The Kohn–Sham equation is solved using the full external potential: $V_{\text{ion}}(\mathbf{r}) - e\vec{\mathcal{E}} \cdot \vec{r}$. For quantities such as polarizability, which are derivatives of the total energy, one can compute the energy at a few field values and differentiate numerically. Real-space methods are suitable for these calculations, since the position operator $\vec{r}$ for a confined system such as a cluster is not ill-defined, as would be the case for a crystal.

Determining the polarizability of a cluster (or molecule) remains difficult owing to the need for a complete basis in the presence of an electric field. To be more precise, if we optimize a basis for $V_{\mathrm{ion}}(\vec{r})$, will the basis also be optimized in the presence of an electric field: $V_{\mathrm{ion}}(\mathbf{r}) - e\vec{\mathcal{E}} \cdot \vec{r}$? Possibly not, and certainly not for a larger field. The basis in the presence of a field is sometimes augmented by "polarization functions." Augmenting the basis can significantly complicate the process. We really want to avoid optimizing the basis each time we apply a different external field or make a change in the structure of the cluster.

Real-space and plane wave methods are not biased as are localized bases and they do not depend on atomic positions. The absence of a "bias" means that angular components of the wave functions are on equal footing. A real-space method implemented with a uniform grid possesses an "isotropic" environment with respect to the applied field. Moreover, the convergence of an electronic property can be easily checked by changing the grid size, especially for a cubic system where only one parameter characterizes the grid.

Example 12.1 ***Polarizabilities in Sodium Clusters*** To illustrate this procedure, we consider the polarizability of sodium clusters, which serve as archetypal simple metal clusters. Measurements for Na cluster polarizabilities as a function of cluster size are illustrated in Figure 12.5 [10]. Three main features are present in the experimental data: (i) Overall, the polarizability per atom gradually decreases from its atomic value toward its bulk value. (ii) This gradual decrease is punctuated by significant "dips" at "magic" atom numbers of 2, 8, and 18, corresponding to closed electronic shells of *s*-, *p*-, and *d*-like orbital character, respectively. (iii) Some residual oscillations occur. Such fluctuations are attributed to small energy differences between nearly isomeric structures.

Early calculations consistently underestimated the experimentally observed value by ~15–20% [11]. Some initially attributed this discrepancy to a failure of density functional theory, which is a disconcerting conclusion as Na is probably the element closest to replicating a homogeneous electron gas.

There is a clue in Figure 12.5 as to where the problem might reside and it is not related to density functional theory. We note that the polarizability of the atom agrees pretty well with experiment. This suggests that density functional theory is not the problem. So, why would the polarizability of the atom, but not a cluster, be correctly predicted by density functional theory? The atom is special in that it lacks any atomic structure. This observation led investigators to ask why the atomic structure is important and why it might not be correctly described in the computations.

An obvious issue, especially one in retrospect, concerns the role of temperature. The experimental values were not measured at absolute zero. Rather, the clusters were "hot," with an effective temperature near the melting point.

Suppose the clusters experience a thermal expansion, which might be important in determining the structure of a cluster. Classically, the polarizability of a sphere scales as the volume of the sphere. Consequently, a thermal expansion of ~ 5% would account for the difference between theory and experiment. In Figure 12.5, calculations for the polarizability of Na clusters are illustrated with

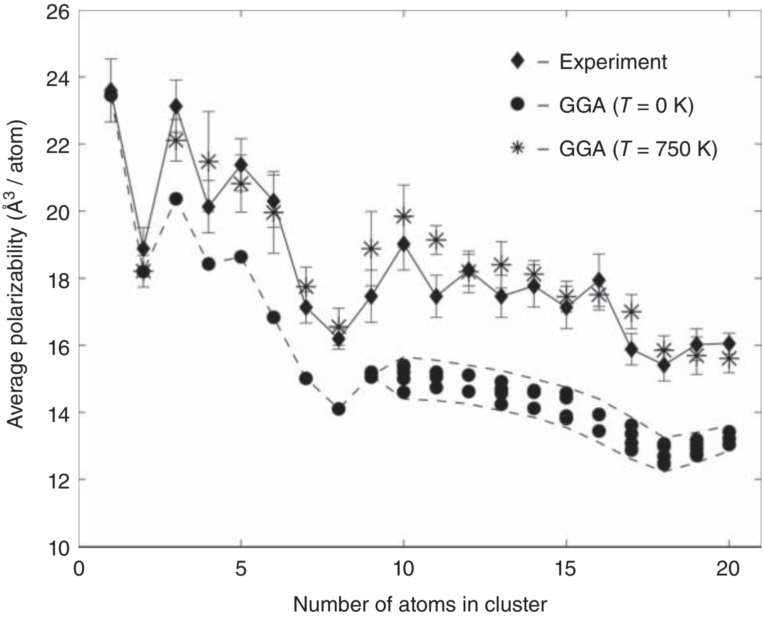


Figure 12.5 Polarizability of Na clusters. The experiment is from Knight and coworker [10]. Two theoretical simulations [4] are illustrated. One is at $T = 0K$, and the other is at a high temperature, $T = 750K$, as indicated. GGA refers to the type of density functional theory used in the computation.

and without temperature. The zero temperature results were calculated based on the low-energy structures of Figures 12.2 and 12.3.

The high-temperature results were obtained by averaging over geometries obtained by molecular dynamics with a heat bath temperature fixed at 750 K.

A molecular dynamics simulation within Langevin dynamics can be performed to generate ensemble averages for a cluster within the heat bath. This can be justified by considering a cluster within a buffer gas. The atoms in the clusters would receive random "kicks" by atoms in the buffer gas, which might be similar to what one would expect from Langevin dynamics. Of course, a real buffer gas might not have the detailed characteristics of the heat bath.

The zero temperature polarizabilities mimic the general shape of the experimental curve, but they are consistently lower than the experimental values. If temperature is included, the theoretical polarizabilities increase due to cluster expansion and distortion, and can be brought into registry with experiment with a suitable choice of temperature. The inclusion of thermal expansion reconciles the discrepancy attributed to the failure of density functional theory.

12.3.2 The Optical Properties of Clusters

We know that the interaction of light with clusters is a useful tool for exploring the electronic and structural properties of clusters. It can be used to distinguish one possible structure from another.

Understanding the interaction of light with clusters can be handled in a manner similar to that for molecules, e.g. using time-dependent density functional theory as in Section 11.5.1. The only complication for clusters centers on an issue of structure, which is almost always known for molecules.

We again consider sodium clusters. The optical properties of these clusters have been calculated with time-dependent density functional theory. Na clusters should be good candidates for this method, given their "free electron" nature.

Time-dependent density functional theory often works well for simple metals. Accurate experimental measurements of the absorption spectra for Na clusters are available, and some of these measurements date back to the first half of the twentieth century. As for the case of molecules, the the unoccupied electron states of clusters can be very sensitive to the boundary conditions as they are often quite extended. For real-space calculations, this situation requires calculations that are performed within a relatively large boundary domain. One can assess the situation by changing the domain size and repeating the calculation.

The calculated absorption spectrum for some small closed shell Na clusters is shown in Figure 12.6 along with experiment. In addition, we illustrate the spectrum generated by considering transitions between the Kohn–Sham eigenvalues. The agreement between the spectra calculated from the time-dependent local density approximation and the measured spectra is remarkable, especially when contrasted with the Kohn–Sham spectra.

The time-dependent local density approximation correctly reproduces the experimental spectral shape, and the calculated peak positions agree with experiment within 0.1–0.2 eV. One of the most intriguing differences between the calculated spectra is the collective nature of the time-dependent local density approximation. A simple shift of the Kohn–Sham spectra would not reproduce the experimental spectra. This is particularly the case for Na_4 where distinct new spectral features occur. As one might expect, the accuracy of the time-dependent local density approximation for the Na clusters is not "typical." While the approximation works well here, it does not work so well for clusters in general, especially as the size of the system increases.

The time-dependent local density approximation cannot be better than the underlying theory. In particular, since density functional theory often fails to localize the electron–hole pair created by an optical excitation, the method becomes flawed for large clusters, especially when the electron–hole pair interactions are large. For small systems, the physical size of the system confines the interactions, and the results are better.

12.4 The Role of Temperature on Excited-state Properties

Size is not the only characteristic that can have a strong effect on the properties of a cluster. As we found for polarizabilities, temperature can also play an important role in interpreting optical spectra.

As an example, we focus again on Na clusters. The optical absorption of such clusters can exhibit an unexpectedly large temperature dependence. How does one account for this? Temperature effects are often simulated by finite broadening of the optical absorption spectra by Gaussians or similar functions such as Lorentzian (see Eq. 10.43). This is the case in Figure 12.6 where we broadened the lines by convoluting the spectra with a Gaussian function.

Is there a justification for broadening the line shape? Often a simple broadening of the spectrum appears consistent with experiment. The spectral line shape, even for a well-defined transition, is never an infinitely sharp line.

There are several reasons for this. We discussed one of them earlier. For a line to be infinitely narrow, we have to make the measurement over an infinite time. In general, that is not the problem. Even if an infinitely narrow spectral line were incident on a spectrometer, the measured line would not be infinitely narrow as spectrometers do not have infinite resolution.

Still, one would like to have a better understanding of temperature and how it might broaden the spectrum by incorporating it into calculations for the absorption spectra. In principle, this could be done by considering a finite temperature simulation and observing its effect on the optical spectra of clusters. For example, one could put a molecule in a heat bath at some given temperature and collect an ensemble average. By changing the temperature of the heat bath, it should be possible to monitor the role of the temperature.

There are notable issues with this procedure. In particular, suppose we compute excited-state spectra using time-dependent density functional theory for each step in a molecular dynamics simulation. This will work. However, time-dependent density functional calculations are not cheap. One would like to sample hundreds if not thousands of spectra on a large ensemble of molecules. This process is tough, but it is not beyond the pale of modern computers. Provided the calculations can be done quickly a decent-sized sample can be obtained.

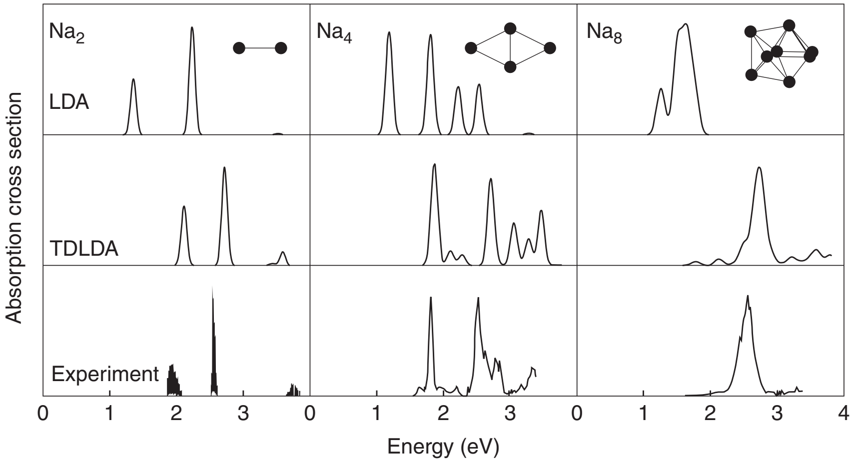


Figure 12.6 Calculated [12] and experimental [13], [14] absorption spectra for Na_2, Na_4, and Na_8. The top panel shows a local density approximation to the spectrum using Kohn–Sham eigenvalues. The middle panel shows a time-dependent local density approximation calculation. The bottom panel shows experiment.

We illustrate this technique by comparing to experimental work performed on small sodium clusters, which have been ionized and mass selected [15].

Example 12.2 ***Finite Temperature Simulations for the Optical Properties of Sodium Clusters*** We place a sodium cluster in a thermal bath and run numerical experiments. Initially the cluster is allowed to thermalize with the heat bath. Then samples are taken at given time intervals. An ensemble average of the optical spectra will usually converge fairly quickly for these clusters and a large number of sampled points are not needed. In Figure 12.7, we illustrate results for a Na_7^+ cluster.

Two temperature regimes are illustrated. At a low temperature, $T = 35$ K, a Gaussian is used to broaden the spectrum so that it agrees with experiment. This broadening is assumed to be instrument dependent and not temperature dependent. The line shape, which arises from the instrument, need not be Lorentzian. At the high temperature, $T = 650$ K, we keep this broadening, but add in thermal broadening. In this way the broadening of the spectra from the low temperature to the high temperature calculation is due only to the time-averaging of different configurations in the simulation. This simple approach works reasonably well, and the broadened spectra are consistent with experiment. However, this procedure remains computationally intensive.

12.4.1 Magnetic Clusters of Iron

So far, our work on clusters has focused on non-magnetic systems. Magnetic clusters can be handled within pseudopotential-density functional theory, although the computations are more difficult as a new degree of freedom is introduced.

The existence of spontaneous magnetization in metallic systems is an intriguing problem because of the extensive technological applications of magnetic phenomena and the incomplete theory of its fundamental mechanisms.

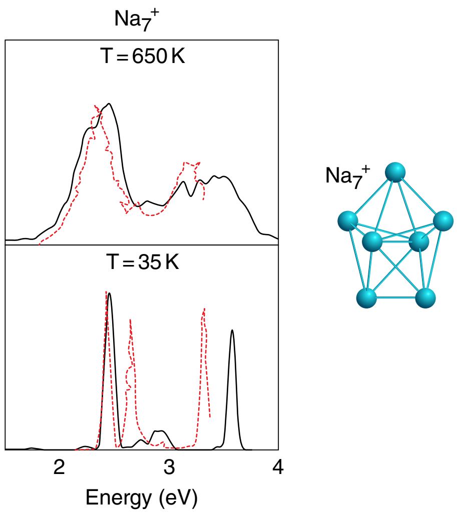


Figure 12.7 Absorption spectra for Na_7^+ clusters at 35 and 650 K. The solid lines are calculated spectra [15] while the dashed curves are from experiment [16].

We focus on the quintessential magnetic clusters – clusters of iron atoms. Spontaneous magnetization is one of the most remarkable properties of iron and materials containing iron, so much that it gave a name to this specific type of magnetization: ferromagnetism from the Latin word for iron (ferrum). One of the more intriguing questions about iron clusters is how they evolve into magnetic crystals – the emergence of magnetism.

At the atomic level, ferromagnetism is associated with partially filled 3d orbitals. In solids, ferromagnetism may be understood in terms of the itinerant electron model, which assumes partial delocalization of the 3d orbitals. In clusters of iron atoms, delocalization is weaker owing to the presence of a surface, whose shape affects the magnetic properties of the cluster. And, because of their small size, iron clusters containing a few tens to hundreds of atoms are superparamagnetic: the entire cluster serves as a single magnetic domain, with no internal grain boundaries. Consequently, small iron clusters have strong magnetic moments, but exhibit no hysteresis.

One difficulty with computing the properties of iron clusters is our old nemesis: The structure of iron clusters is simply not well known. First-principles and model calculations imply that clusters with up to about 20 atoms assume a variety of distinct shapes in their lowest energy configuration [17], [18]. For larger clusters, there is indication in experiment for a stable body-centered cubic structure, which is identical to ferromagnetic bulk iron [19].

The evolution of magnetic moment as a function of cluster size has attracted considerable attention from researchers in the field. A key question to be resolved is, *what drives the suppression of magnetic moment as clusters grow in size?*

In an iron atom, the permanent magnetic moment arises from the exchange splitting: the $3d_\uparrow$ orbitals (majority spin) are lower in energy and so completely occupied with five electrons. The $3d_\downarrow$ orbitals (minority spin) are partially occupied with one electron, resulting in a magnetic moment of $4\mu_\mathrm{B}$, μ_B being the Bohr magneton. (Recall from Section 2.3:$\mu_\mathrm{B} = e\hbar/2mc$.)

When atoms assemble in a crystal, their atomic orbitals hybridize and form energy bands: the 4s hybridized orbitals create a wide band that remains partially filled, in contrast with the completely filled 4s orbital in the atom. The $3d_\downarrow$ and $3d_\uparrow$ orbitals create narrower bands. Orbital hybridization together with the different bandwidths of the various 3d and 4s bands result in weaker magnetization, equivalent to 2.2 μ_B/atom in bulk iron.

In atomic clusters, orbital hybridization is not as strong because atoms on the surface of the cluster have fewer neighbors. A theoretical analysis in clusters and thin slabs indicates that the dependence of the magnetic moment with the effective coordination number is approximately linear and increases with coordination [20].

The suppression of the magnetic moment from orbital hybridization is not isotropic, if we consider a layer of atoms. For instance, the 3d orbitals oriented in the plane of atoms will hybridize more effectively than orbitals oriented normal to the plane. As a consequence, clusters with faceted surfaces are expected to have magnetic properties different from clusters with irregular surfaces, even if they have the same effective coordination number. This effect is likely responsible for a non-monotonic suppression of magnetic moment as a function of cluster size.

Example 12.3 ***Evolution of Magnetism in Iron Clusters*** Suppose we examine iron clusters containing up to several hundred atoms. Can we replicate the evolution of magnetism in iron clusters?

An accurate description of the electronic and magnetic structure of iron clusters is more difficult to compute than for simple metal clusters. The existence of a magnetic moment means that an additional degree of freedom enters the problem. If we take the coordinates of an electron plus an up or down spin, we double the size of the problem. Since our machinery for solving the electronic structure problem does not scale linearly but rather as the cube of the matrix size, doubling the matrix size roughly means an order of magnitude increase in terms of computational operations. Also, we should consider "non-collinear magnetism" and associate a magnetic vector at every point in space, which makes the problem even worse. Fortunately, a collinear description is all right for the case at hand owing to the high symmetry of the clusters considered.

Another complicating issue is the relatively localized nature of the 3d electronic states, which requires a fine grid to obtain a fully converged solution. For iron the grid spacing might be less than half of that for silicon, i.e. 0.3 a.u. (Fe) versus 0.7 a.u. (Si). A fine grid results in a much larger Hamiltonian matrix and a corresponding increase in the computational load. While we can accurately compute the properties of nanocrystals of silicon with over thousands of atoms, nanocrystals of iron of this size are problematic without additional computing power.

Things could get much worse. The geometry of the iron clusters introduces a number of degrees of freedom, as we have stated numerous times. It is not currently possible to determine the *definitive* ground state for systems with dozens of atoms. However, in many cases it is not necessary to know the ground state. We are more interested in determining what structures are "reasonable" and representative of the observed ensemble, i.e. if two structures are within a few millielectron volts, they are not distinguishable.

In our example, we build topologically distinct clusters, e.g. clusters of both icosahedral and body-centered cubic symmetry. To investigate the role of surface faceting, we constructed clusters with faceted and non-faceted surfaces. Faceted clusters are constructed by adding successive atomic layers around a nucleation point. Small faceted icosahedral clusters exist with special sizes: 13, 55, 147, and 309 [20]. Faceted body-centered clusters are constructed with the same local coordination as the bulk crystal.

Suppose we consider two families of cubic clusters: "atom centered" or "bridge centered." These terms mean what they imply. An atom-centered cluster has a nucleation point at an atom site. A bridge-centered cluster has a nucleation point between two neighboring atoms. Non-faceted clusters are built by adding shells of atoms around a nucleation point so that their distance to the nucleation point is less than a specified value. Non-faceted clusters usually have narrow steps over otherwise planar surfaces and the overall shape is almost spherical. By construction, non-faceted clusters have well-defined symmetries. Clusters constructed

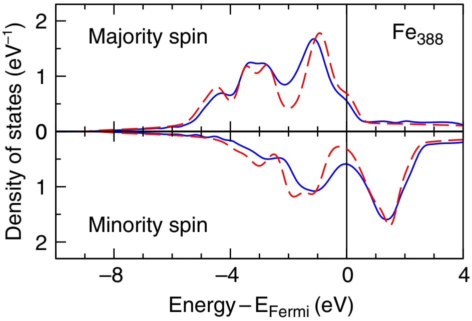


Figure 12.8 Density of states for the iron Fe_{388} body-centered cubic-like cluster. Majority spin (upper panel) and minority spin (lower panel). For reference, the density of states in bulk iron is shown in dashed lines. The Fermi energy is chosen as energy reference.

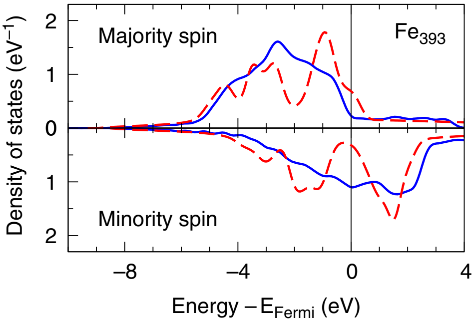


Figure 12.9 Density of states for the iron Fe_{393} icosahedral-like cluster. Majority spin (upper panel) and minority spin (lower panel). For reference, the density of states in bulk iron is shown in dashed lines. The Fermi energy is chosen as energy reference.

in this manner show low tension on the surface, making surface reconstruction less likely.

As clusters grow in size, their properties approach the properties of bulk iron as is apparent in Figure 12.8. Here we display the density of states for Fe_{388}, which assumes body-centered cubic-like coordination. If we compare the bulk iron density of states to the cluster, the results are quite similar. In addition, the cohesive energy in this computation is only 77 meV/atom lower than in the bulk [20]. Such computations suggest that quantum confinement effects will be predominantly observed in clusters smaller than Fe_{388}.

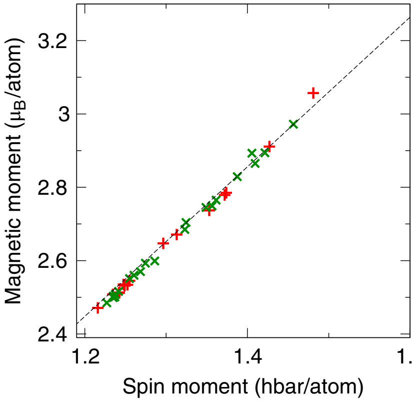
Figure 12.9 shows the density of states for Fe_{393}, which belongs to the icosahedral family. This cluster has a "nondescript" density of states, i.e. one without much structure. One can attribute this difference to the icosahedral-like arrangement of atoms. The overall dispersion of the 3d peak (4 eV for $3d_{\uparrow}$ and 6 eV for $ed_{\downarrow}$) is similar in both density of states.

The key issue for the example at hand is to determine how the magnetic moments for iron clusters evolve. The magnetic moment is calculated as the expectation value of the total angular momentum:

$$M = \frac{\mu_B}{\hbar}[g_s\langle S_z\rangle + \langle L_z\rangle] = \mu_B\left[\frac{g_s}{2}(n_{\uparrow} - n_{\downarrow}) + \frac{1}{\hbar}\langle L_z\rangle\right] \tag{12.15}$$

where $g_s = 2$ is the electron gyromagnetic ratio. Figure 12.10 illustrates the approximately linear dependence between the magnetic moment and spin moment, $\langle S_z\rangle$, throughout the whole size range.

Figure 12.10 Magnetic moment versus spin moment calculated for the atom-centered body-centered cubic ("plus" signs) and bridge-centered body-centered cubic (crosses) iron clusters. The approximate ratio is $M/\langle S_z \rangle = g_{eff} = 2.04\mu_B/\hbar$.



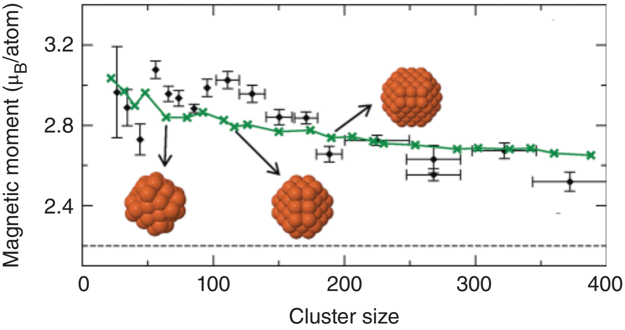


Figure 12.11 Calculated magnetic moments for clusters in a body-centered-like structure. Some of the faceted and non-faceted clusters are depicted next to their corresponding data points. The dashed line indicates the value of magnetic moment per atom in bulk iron.

This results in an effective gyromagnetic ratio $g_{\text{eff}} = 2.04\mu_B/\hbar$, which is somewhat smaller than the gyromagnetic ratio in bulk body-centered cubic iron, $2.09\mu_B/\hbar$. This is most likely due to an underestimation of the orbital contribution, $\langle L_z \rangle$. In the absence of an external magnetic field, orbital magnetization arises from the spin–orbit interaction, which is included in the theory as a model potential [21]:

$$V_{\text{so}} = -\xi \mathbf{L} \cdot \mathbf{S} \tag{12.16}$$

where $\xi = 80$ meV/$\hbar^2$.

Figure 12.11 shows the magnetic moment of some body-centered cubic generated clusters. Figure 12.12 shows the moments of some icosahedral like clusters. Experimental data [19] are also shown in both figures. The suppression of magnetic moment as a function of size is readily observed. Also, clusters with faceted surfaces are predicted to have magnetic moments lower than other clusters with similar sizes. This is attributed to more effective hybridization of d orbitals along the plane of the facets.

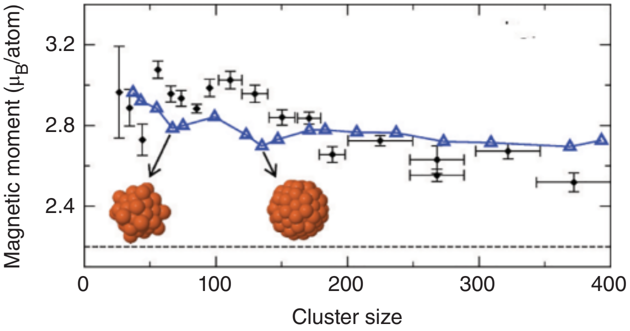


Figure 12.12 Calculated magnetic moments for clusters in an icosahedral-like structure. Some of the faceted and non-faceted clusters are depicted next to their corresponding data points. The dashed line indicates the value of magnetic moment per atom in bulk iron.

The non-monotonic behavior of the measured magnetic moment with size cluster can be attributed to the shape of the surface. Under this assumption, islands of low magnetic moment are associated to clusters with faceted surfaces.

References

1 Schlick, T. (2002). *Molecular Modeling and Simulation*. Springer-Verlag.

2 Binggeli, N., Martins, J.L., and Chelikowsky, J.R. (1992). Simulation of Si clusters via Langevin molecular dynamics with quantum forces. *Physical Review Letters* 68: 2956.

3 Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2007). *Numerical Recipes: The Art of Scientific Computing*, 3e. Cambridge University Press.

4 Kronik, L., Vasiliev, I., and Chelikowsky, J.R. (2000). *Ab initio* calculations for the structure and temperature effects on the polarizabilities of Na_n ($n \leq 20$) clusters. *Physical Review B* 62: 9992.

5 Deaven, D.M. and Ho, K.M. (1995). Molecular geometry optimization with a genetic algorithm. *Physical Review Letters* 75: 288.

6 Wales, D.J. and Doye, J.P.K. (1997). Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry* 101: 5111.

7 Jackson, K.A., Horoi, M., Chaudhuri, I. et al. (2004). Unraveling the shape transformation in silicon clusters. *Physical Review Letters* 93: 013401.

8 Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. *Proceedings of IEEE International Conference on Neural Networks* 4: 1942.

9 Poli, R., Kennedy, J., and Blackwell, T. (2007). Particle swarm optimization. *Swarm Intelligence* 1: 33.

10 Cohen, M.L. and Knight, W.D. (1990). The physics of metal clusters. *Physics Today* 43: 42.

11 Kümmel, S., Akola, J., and Mannien, M. (2000). Thermal expansion in small metal clusters and its impact on the electric polarizability. *Physical Review Letters* 84: 3827.

12 Vasiliev, I., ”Oğüt, S., and Chelikowsky, J.R. (2002). First principles density functional calculations for optical spectra of clusters and nanocrystals. *Physical Review B* 65: 115416.

13 Wang, C.R.C., Pollack, S., Cameron, D., and Kappes, M.M. (1990). Molecular excited states versus collective electronic oscillations: optical absorption probes of Na_4 and Na_8. *Chemical Physics Letters* 93: 3787.

14 Fredrickson, W.R. and Watson, W.W. (1927). The sodium and potassium absorption bands. *Physical Review* 30: 429.

15 Lopez del Puerto, M., Tiago, M.L., and Chelikowsky, J.R. (2007). *Ab initio* calculation of temperature effects in the optical response of open-shell sodium clusters. *Journal of Chemical Physics* 27: 144311.

16 Ellert, C., Schmidt, M., and Schmitt, C. (1995). Temperature dependence of the optical response of small, open shell sodium clusters. *Physical Review Letters* 75: 1731.

17 Pastor, G.M., Dorantes-Dávila, J., and Bennemann, K.H. (1989). Size and structural dependence of the magnetic properties of small 3d-transition-metal clusters. *Physical Review B* 40: 7642.

18 Diéguez, O., Alemany, M.G.G., Rey, C. et al. (2001). Density-functional calculations of the structures, binding energies, and magnetic moments of Fe clusters with 2 to 17 atoms. *Physical Review B* 63: 205407.

19 Billas, I.M.L., Beckner, J.A., Châtelain, A., and de Heer, W.A. (1993). Magnetic moments of iron clusters with 25 to 700 atoms and their dependence on temperature. *Physical Review Letters* 71: 4067.

20 Tiago, M.L., Zhou, Y., Alemany, M.G.G. et al. (2006). Evolution of magnetism in iron from the atom to the bulk. *Physical Review Letters* 97: 147201.

21 Šipr, O., Košuth, M., and Ebert, H. (2004). Magnetic structure of free iron clusters compared to iron crystal surfaces. Physical Review B 70: 174423.

13

Nanocrystals

> *When we get to the very, very small world – say circuits of seven atoms – we have a lot of new things that would happen that represent completely new opportunities for design. Atoms on a small scale behave like nothing on a large scale, for they satisfy the laws of quantum mechanics. So, as we go down and fiddle around with the atoms down there, we are working with different laws, and we can expect to do different things. We can manufacture in different ways. We can use, not just circuits, but some system involving the quantized energy levels, or the interactions of quantized spins, etc.*
>
> Richard Feynman

Richard Feynman made the comments cited above in 1959 at Caltech. His lecture was entitled “There is Plenty of Room at the Bottom.” Feynman speculated that atomic manipulation would lead to efficient computers and information storage devices. Some claim that Feynman’s remarks inspired much of the nanoscience and nanotechnology work performed today. However, this claim is not substantiated by a study of the literature [1]. To the contrary, Feynman’s lecture did not have a great impact at the time or in the immediate following years. Nonetheless, his talk was prescient.

In Chapter 12, we defined what a cluster is. We do the same for a “nanocrystal.” A nanocrystal is not really the same as a “big” cluster with a crystalline structure. An operational definition for a “nanocrystal” can be constructed by considering how properties change with size. Consider a macroscopic piece of silicon, say, something you could hold in your hand. Suppose we examine the optical properties of this piece of silicon. Cutting the piece in two does nothing to change it optical characteristics. For example, a 10 cm^3 sample is no different in its optical properties from a 10 mm^3 sample. Both look like silicon. Properties that do not change with size or mass are called “intensive.” An optical gap is an intensive property as is the density of the material. Now suppose we keep cutting the original piece of silicon into smaller and smaller size. As we go to smaller and smaller length scales, e.g. sizes below 100 nm, something unusual happens. The optical properties begin to change with size and cease to be intensive. At this point, we have a “nanocrystal.” A working definition of a nanocrystal might be a crystalline fragment at a length scale wherein bulk intensive properties are no longer independent of size.

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

A key aspect of nano-related phenomena is the role of "quantum confinement." At small length scales, quantum mechanical properties become important as Feynman suggested. We can use elementary arguments to determine at what length scale quantum mechanics becomes important.

We start with the *uncertainty principle*. This principle states that the uncertainty of a particle's momentum, Δp, and its position, Δx, must be such that at best $\Delta p \Delta x \sim \hbar$, where h is Planck's constant divided by 2π. Suppose we consider one dimension and put a particle in a box of size Δx. We know then that the momentum must be $p \sim \Delta p = \hbar/\Delta x$. The kinetic energy of a particle in this box should scale as $\sim 1/\Delta x^2$ and we expect that the kinetic energy will rapidly increase for small values of Δx. In a real particle, a three-dimensional one, the idea is the same. Confining an electron to a small physical volume requires energy.

If the confinement energy becomes comparable to the characteristic energy of the crystal such as its band gap, then its physical properties will change. In general, once the confining dimension approaches the delocalization length of an electron in a solid, quantum phenomena will happen. The delocalization length can vary from solid to solid and is correlated with the wave properties of the electron. For example, the de Broglie wavelength of an electron is given by $\lambda = h/p$. If λ is on the order of Δx, then we expect confinement and quantum effects to become important.

For a simple metal, we can estimate the momentum from a free electron gas: $p = \hbar k_\mathrm{f}$. k_f is the wave vector such that the Fermi energy, $E_\mathrm{f} = \hbar^2 k_\mathrm{f}^2/2m$. This yields $k_\mathrm{f} = (3\pi^2 n)^{1/3}$ where n is the electron density. The value of k_f for a typical simple metal such as Na is $\sim 10^{10}\mathrm{m}^{-1}$. This would give a value of $\lambda = h/p = 2\pi/k_\mathrm{f}$ or roughly $\sim$1 nm. This characteristic "nano" scale gives its name to nanostructures or nanocrystals for the case at hand.

To observe the role of quantum confinement in real materials, scientists and engineers need to construct materials routinely on the nanometer scale. Nanocrystals of many materials can be made, although sometimes it is difficult to determine whether they are crystalline. Nanocrystals provide science a unique opportunity to study the properties at nanometer scales and to reveal the fundamental physics (read – "quantum effects") occurring at reduced dimensionality. From a "technological" point of view, nanostructures are currently the building blocks in electronic devices, e.g. the smallest nominal length scale in a modern CPU chip is less than 100 nm. As this size shrinks, electronic device engineers are becoming "quantum" mechanics.

At such sufficiently small length scales, the band structure of a material may no longer appear to be quasi-continuous. Instead of a "band," the electronic energy levels are quantum energy levels, which change with the size of the system. A deeper understanding of the physics of confinement is necessary to exploit these new laws for the development of nano optical, magnetic, and electronic device applications. This understanding can be obtained by utilizing the same approaches as in the previous chapters for clusters or molecules. But, the problem is now harder, much harder. In an odd twist of terminology, nano scale problems involve "large-scale" computations. A nanocrystal can often contain thousands, if not millions, of atoms – much more computationally intensive than clusters containing a few dozen atoms.

13.1 Semiconductor Nanocrystals: Silicon

Our nanocrystal discussion begins with silicon, which, along with sodium, is one of the usual suspects in condensed matter. Just as atomic physicists often consider a hydrogen atom as a testbed, solid-state physicists use silicon [2].

Why silicon? The answer is simple: silicon is the most important technological material of our time. Scientists and engineers have performed numerous measurements on its chemical and physical properties. The number of scientific papers on silicon exceeds several hundred thousands. The vastness of the silicon literature makes it appealing as a testbed for new theory and experiment.

13.1.1 Intrinsic Properties

13.1.1.1 Electronic Properties

If we wish to examine the intrinsic properties of silicon nanocrystal, we need to deal with how to terminate the nanocrystal. How do we handle the surface, which at the nanoscale becomes increasingly important? If one were to construct a nanocrystal by simply extracting a bulk fragment, atoms at the surface will be under-coordinated. In the case of silicon, unsaturated bonds will result in defect states. The existence of such states complicates things. They will produce energy levels within the band gap and the nanocrystal will become a metal. This is not what we want to study theoretically nor do we want this if we were doing experiments. In experiment, the dangling bonds in a nanocrystal are saturated or passivated, i.e. rendered chemically inert. Passivation can be accomplished by coating the surface with polymers, surfactants, or hydrogen atoms. Passivation can also be accomplished by embedding the nanocrystal in a medium. Ideally, a passivated nanocrystal should not exhibit extrinsic states in the band gap, and ideally maintain a bulk-like structure right up to the surface.

Perhaps the easiest passivation mechanism for computational studies is to cap surface dangling bonds with hydrogen-like atoms. The only purpose of these capping atoms is to terminate the nanostructure in a gentle manner, both structurally and electronically. Hydrogen atoms on a silicon nanocrystal, i.e. a small crystal of silicon, form Si—H bonds. The energy levels corresponding to these Si—H states do not reside with the silicon band gap, but rather a few volts below the gap. Moreover, the Si—H bonds and surrounding Si—Si bonds can be structurally relaxed so that the surface is not under strain.

A simple model for constructing a silicon nanocrystal is to consider all the atoms within a given radius of a particular atom. Hydrogen atoms at the surface are then used to saturate any broken bonds.

Nanocrystals with several thousands of atoms can be examined using the algorithms outlined in Chapter 9. Computational issues center on solving the Kohn–Sham problem for a large number of electronic states. The problem can be made somewhat simpler by using symmetry and by the use of algorithms that allow an eigenvalue spectrum to be solved with a specified energy window. The problem remains difficult and not many calculations are done at this time for systems with more than tens of thousands of atoms. As computing power increases, the number of atoms will also increase. (Not too many years ago a hundred atoms of silicon was considered a large system. So, one can expect that nanocrystals with hundreds of thousands of atoms are not out of the question in the near future.)

As an example, we will focus on $Si_{9041}H_{1860}$. This nanocrystal is larger than the smallest Si nanocrystal made in the laboratory. Since the inception of nanoscience, experimentalists have worked to synthesize and characterize smaller and smaller nanostructures while theorists have worked to compute properties for larger and larger nanostructures. Experimentalists initially could not control the synthesis of nanocrystals smaller than a few thousand atoms and theorists initially could not compute properties for nanocrystals larger than about a hundred atoms. Sometime in the early part of the twentieth century, the two endeavors succeeded in describing the same-sized nanostructures, after which direct comparisons between experiment and theory became feasible.

A ball and stick model of a typical nanocrystal is shown in Figure 13.1. A solution of the Kohn–Sham equations yields the distribution of eigenvalues. For sufficiently large nanocrystals, one expects the distribution to approach that of crystalline silicon, i.e. the distribution of states should approach that of the crystalline "density of states." Is a 7 nm nanocrystal of Si of sufficient size to have a spectrum that looks like bulk silicon? The answer is "yes." A comparison is made in Figure 13.2. The "density of states" for the eigenvalue spectrum for the nanocrystal shares a similar structure as the crystalline density of state.

The only notable differences between the crystal and nanocrystal densities of states occur with the presence of the Si—H bonds. These states occur about 6 eV below the highest occupied states (taken as our zero reference). Another difference not particularly evident in the figure is the discrete nature of the eigenvalue spectrum for the nanocrystal. We have 38 024 electrons in our nanocrystal, which occupy 19 012 states. The energy of the occupied eigenvalues span about 12 eV. If uniformly distributed, we would have an average spacing of 0.006 eV, which is below the energy associated with room temperature (0.025 eV). Eigenvalues so distributed are quasi-continuous.

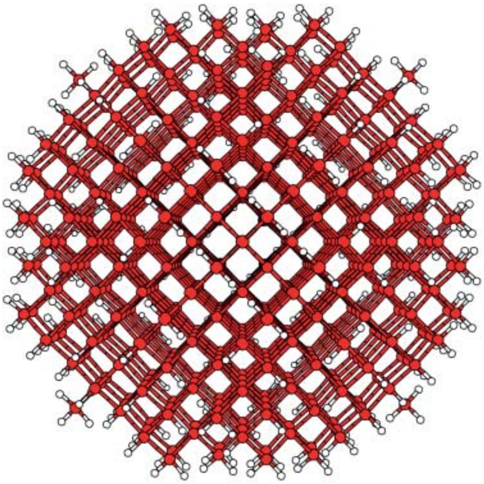


Figure 13.1 A ball and stick model of a hydrogenated silicon nanocrystal. The interior consists of a diamond fragment. The surface of the fragment is capped with hydrogen atoms.

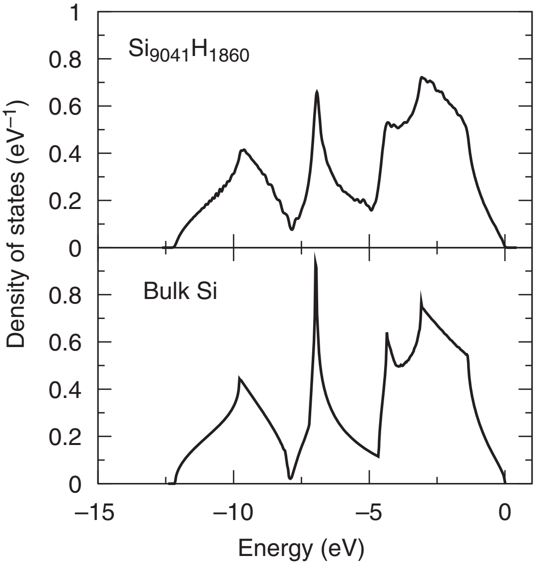


Figure 13.2 The eigenvalue density of states for the $Si_{9041}H_{1860}$ nanocrystal (top panel) and the electronic density of states for crystalline silicon (bottom panel). Only occupied states are considered. The highest occupied state in both panels is taken to be the zero energy reference.

How far is our nanocrystal from the bulk limit? We can try another test. We examine how the energy gap in silicon evolves with size. (We will use "energy gap" and "band gap" interchangeably as the energy gap in a nanocrystal should evolve to a band gap for a sufficiently large nanocrystal.) We keep in mind that the size of a band gap is not likely to be correct when computed within the density functional theory, which is a ground-state theory. While trends in gaps are often correctly given by density functional theory, absolute values are often off by a factor of two or more. Here are some examples. The measured band gap of crystalline silicon is 1.1 eV; the calculated gap is 0.6 eV. The measured gap in silica is roughly 10 eV; the calculated gap is about 5 eV.

There is a bit of mystery here. In Chapter 10, we noted that the ionization energy calculated for atoms was not bad, save a few anomalies associated with spin. In this case, density functional theory worked well.

Why does it fail for large or extended systems? A possible explanation centers on the failure of density functional theory to produce localized excitations. When light is absorbed in a semiconductor or insulator, an electron–hole pair is created. In silicon, the excitation is localized to roughly 5 nm. Density functional theory may not include the correct interactions of the electron–hole pair and the excitation is not properly localized. Accordingly, the electron–hole pair can be localized by the physical size of the system, e.g. the electron–hole excitation in an atom is localized by the physical size of the atom.

Within density functional theory, one can construct a well-defined measure for creating a *non-interacting* electron–hole pair. First, we find the energy to remove an electron from the system and create a hole. For a semiconductor, this means we remove an electron from the highest occupied state and move it to the vacuum level. This energy is called the ionization potential (IP), I. The energy to add an electron to the lowest unoccupied state is called the electron affinity (EA), A.

These energies are defined as follows.

$$\begin{aligned} I &= E(N-1) - E(N) \\ A &= E(N) - E(N+1) \end{aligned} \tag{13.1}$$

where N is the total number of electrons in the system. In principle, the affinity and ionization energies are ground-state properties and, if the correct functional were known, these quantities would be accurately predicted by density functional theory.

We examine the scaling of the IP and the EA by assuming a simple power law behavior and fitting such a form to the calculated values (shown in Figure 13.3):

$$\begin{aligned} I(D) &= I_\infty + A/D^\alpha \\ A(D) &= A_\infty + B/D^\beta \end{aligned} \tag{13.2}$$

where D is the dot diameter. A fit of these quantities results in $I_\infty = 4.5$ eV, $A_\infty = 3.9$ eV, $\alpha = 1.1$, and $\beta = 1.08$. The fit gives a quasi-particle gap of $E_{qp}(D \to \infty) = I_\infty - A_\infty =$ 0.6 eV in the limit of an infinitely large dot. This value is in good agreement with the gap found for crystalline silicon using the local density approximation [3]. The gap is not in good agreement with experiment owing to the failure of the local density approximation to describe band gaps of bulk semiconductors in general. We learned something interesting here. Even though I and A were computed from ground-state properties for a large system, they do not yield correct properties for the ground state. This situation reflects some fundamental flaw in our choice of functionals, which is not surprising as no functional to date is "perfect."

A key aspect of this example is to show the scaling of the IP and EA for nanocrystals ranging from silane (SiH_4) to hydrogenated silicon systems containing thousands of atoms. We not only verify the limiting value of the quasi-particle gap, we can also ascertain how this limit is reached, i.e. how the

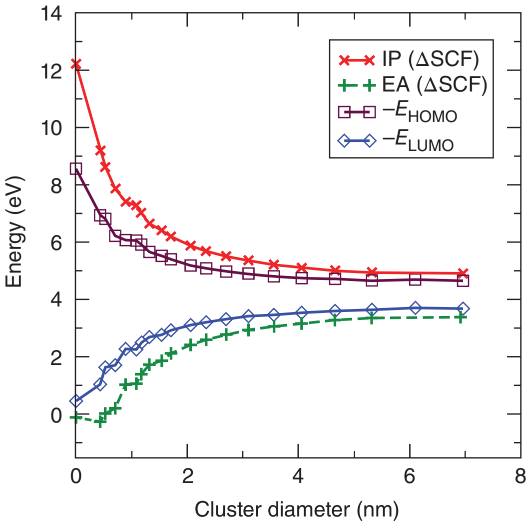


Figure 13.3 Evolution of the ionization potential (IP) and electron affinity (EA) with quantum dot size. Also shown are the eigenvalue levels for the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).

IP and EA scale with the size of the dot and what the relationship is between these quantities and the highest occupied and lowest empty energy levels. In our example, we effectively have gone from the molecular limit to the crystalline limit. We have spanned the entire nanocrystal regime, at least for silicon.

Since values of I and A as calculated from density functional theory are reasonably accurate for atoms and molecules, one can ask how the size of the nanocrystal affects the accuracy of such calculations. Unfortunately, experimental values for the IPs and electron affinities are not known for hydrogenated silicon clusters and nanocrystals, a notable exception being silane, where the EA is negative, which means an electron added to the nanocluster is unbound – not a typical situation [4].

The difference between the IP and the EA can be associated with an energy gap known as the *quasi-particle* gap: $E_{\text{qp}} = I - A$. Formally, this is the energy required to create a non-interacting electron–hole pair.

Some terminology: If the electron and hole are bound together, the excitation is said to form an *exciton*. The concept of an exciton corrects for the failure of a one-electron picture. If we had a full many-electron treatment, there would be no need for these terms.

In crystalline materials, one can make a crude estimate of the exciton energy. Suppose we consider the Coulomb interaction of the electron–hole pair. Recall that a "hole" is a partially filled bond. In crystalline silicon, each atom has four bonds, each shared by its four nearest neighbors. (A technical point: The number of bonds per atom is two, not four, as each bond is shared by a neighboring atom.) If we remove an electron from one of the bonds, there is a charge imbalance with a net charge of +1. An electron will be attracted to this region. If the electron is not too close to the hole, we can treat the interaction as between two point charges. We can write the electron–hole potential as

$$V(r_{\text{eh}}) = \frac{-e^2}{r_{\text{eh}}} \tag{13.3}$$

where r_{eh} is the distance between the electron and the hole.

We can solve for the quantum mechanical energy using this potential as if it were a hydrogen atom. As written, we would make a really big error! Why? First, and foremost, the electron and hole are not in a vacuum. Within the silicon crystal, a point charge is screened by the response of the other electrons. To the lowest order, the electrostatic interactions are screened by a static dielectric constant, ε. Typically, $\varepsilon \sim 10$ for a semiconductor such as silicon. Consequently, the screened potential is reduced by an order of magnitude from the "bare" potential.

We can derive an expression for dielectric constant using a harmonic oscillator description (see Chapter 10 and the discussion following Eq. (10.29)) and show that this is a reasonable number. A simple expression for the dielectric constant is $\varepsilon = 1 + 4\pi\alpha$, where α is the atomic polarizability. This can be generalized to a crystal such as silicon. One finds that

$$\varepsilon(\omega = 0) = 1 + \frac{\omega_{\text{p}}^2}{\omega_0^2} \tag{13.4}$$

ω_p is called the plasmon frequency and ω_0 represents the resonant frequency of a harmonic oscillator. In a semiconductor, the energy $\hbar\omega_\mathrm{p}$ is called the Penn gap. Roughly speaking, it corresponds to the average energy difference between the highest valence band and the lowest conduction band. In semiconductors such as silicon, the plasmon energy, given by $\hbar\omega_\mathrm{p}$, is about 15 eV and the Penn gap is about 5 eV. This yields $\varepsilon \approx 10$.

13.1.1.2 Effective Mass Theory

There is another correction to the electron and hole interaction based on this hydrogenic model. While we now have a rough idea of the potential energy term, what about the kinetic energy? This may seem like an odd question, but holes and electrons are not independent particles and do not behave as such. In particular, what about the "effective masses" of the electron and hole that appear different in solids, or in this case – nanocrystals. The electron and hole move through the crystal as if they had a mass different from each other and from the free electron mass.

To show this result, we can dabble in solid-state physics. It is worth doing as the results are *amazing*. We consider only one energy band in our model, but we could do more if needed. To further simplify the problem, we will assume a one-dimensional problem where the band energy is given by $E(k = |\vec{k}|)$. The crystalline potential that generates this band is given by $V(x)$, where $V(x) = V(x + a)$ is periodic.

$$\left\{ \frac{-\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x) \right\} \psi_k(x) = E(k)\psi_k(x) \tag{13.5}$$

We define the crystalline Hamiltonian as

$$H_0 = \frac{-\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x) \tag{13.6}$$

and will use the fact that $H_0\psi_k = E(k)\psi_k$.

In the presence of an electron–hole interaction, we modify the crystalline potential and seek a solution for the problem:

$$\left\{ \frac{-\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x) + V_\mathrm{eh}(x) \right\} \psi(x) = E\psi_k(x) \tag{13.7}$$

where V_eh represents the potential for the electron–hole interaction, i.e. the Coulomb potential screened by the dielectric constant. The electron–hole potential breaks the translational symmetry of the crystal. The resulting loss of symmetry means Bloch's theorem no longer holds. Before addressing this matter, we need to be creative and exploit the properties of the system without the perturbing potential.

We observe the following:

$$E(k) = \sum_n E(n)\exp(ikna) \tag{13.8}$$

Where did this come from? Recall that $E(k)$ is a periodic function such that $E(k + 2\pi/a) = E(k)$. We are simply expanding $E(k)$ as a Fourier series.

We define an operator by replacing k in $E(k)$ by a derivative $-i\mathrm{d}/\mathrm{d}x$. We consider the action of this operator on ψ_k:

$$E\left(-i\frac{\mathrm{d}}{\mathrm{d}x}\right)\psi_k(x) = \sum_n E(n)\exp\left(na\frac{\mathrm{d}}{\mathrm{d}x}\right)\psi_k(x) \tag{13.9}$$

How do we interpret this? We have an exponential of an operator. The exponential form makes sense only in terms of an expansion:

$$E\left(i\frac{\mathrm{d}}{\mathrm{d}x}\right)\psi_k(x) = \sum_n E(n)\left[1 + na\frac{\mathrm{d}}{\mathrm{d}x} + \frac{(na)^2}{2}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^2 + \cdots\right]\psi_k(x) \tag{13.10}$$

This is a Taylor series expanding around x:

$$\left[1 + na\frac{\mathrm{d}}{\mathrm{d}x} + \frac{(na)^2}{2}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^2 + \cdots\right]\psi_k(x) = \psi_k(x+na) = \exp(ikna)\ \psi_k(x) \tag{13.11}$$

where we used Bloch's theorem: $\psi_k(x+a) = \exp(ika)\ \psi_k(x)$. We rewrite

$$E\left(i\frac{\mathrm{d}}{\mathrm{d}x}\right)\psi_k(x) = \sum_n E(n)\left[1 + na\frac{\mathrm{d}}{\mathrm{d}x} + \frac{(na)^2}{2}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^2 + \cdots\right]\psi_k(x) = E(k)\psi_k(x) \tag{13.12}$$

We have just demonstrated that ψ_k is an eigenfunction of $E(i\frac{\mathrm{d}}{\mathrm{d}x})$ with an eigenvalue of $E(k)$.

We can now reexamine the problem with the electron–hole interaction:

$$\{H_0 + V_{\mathrm{eh}}(x)\}\psi(x) = E\psi(x) \tag{13.13}$$

by writing

$$\psi(x) = \sum_k c(k)\psi_k(x) \tag{13.14}$$

This equation allow us to use ψ_k as a basis to find E. Plugging this in and carrying out some simple substitutions gives us the following:

$$\sum_k c(k)\left\{H_0 + V_{\mathrm{eh}}(x)\right\}\ \psi_k = E\ \psi(x)$$

$$\sum_k c(k)\left\{E(k) + V_{\mathrm{eh}}(x)\right\}\ \psi_k = E\ \psi(x)$$

$$\sum_k c(k)\{E\left(-i\frac{\mathrm{d}}{\mathrm{d}x}\right) + V_{\mathrm{eh}}(x)\}\ \psi_k = E\ \psi(x)$$

$$\{E(-i\frac{\mathrm{d}}{\mathrm{d}x}) + V_{\mathrm{eh}}(x)\}\ \psi = E\ \psi(x)$$

We are almost done.

We finish up by making use of the following. Suppose we are near a band edge; we can expand $E(k)$ around this edge:

$$E(k) = E_c + \frac{\hbar^2 k^2}{2m^*} \tag{13.15}$$

where E_c is the conduction band edge and m^* is an effective mass determined by the curvature of the energy band. We now make the usual substitution $k \rightarrow -i\mathrm{d}/\mathrm{d}x$ and we have the following:

$$\left\{ -\frac{\hbar^2}{2m^*}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V_{\mathrm{eh}}(x) \right\} \psi(x) = (E - E_c)\, \psi(x) \tag{13.16}$$

where we included the electron–hole Coulomb term. In three dimensions, we would get a similar result:

$$\left\{ -\frac{\hbar^2 \nabla^2}{2m^*} - \frac{\mathrm{e}^2}{r_{\mathrm{eh}}} \right\} \psi(\vec{r}) = (E - E_c)\, \psi(\vec{r}) \tag{13.17}$$

What happened here? The problem looks like a Schrödinger equation for a hydrogen atom, but it is different. The Coulomb term is screened by ε and the real electron mass is replaced by an effective mass, m^*. Also, we now measure the energy from the conduction band edge E_c. The role of the crystalline host is to make the mass of the electron look different. This is an amazing result. The electron in the host can still be treated as an electron, but its mass appears different. If we did this for the hole too, we find the effective mass to be $1/m^* = 1/m_e + 1/m_h$. Not surprisingly, this approach is called *effective mass theory.*

If we are given the parameters (ε, m^*, E_c), we can readily write down the answer for the hydrogen-like atom:

$$E = E_c - \frac{m^* \mathrm{e}^4}{2\varepsilon^2 \hbar^2} = E_c - 13.6eV\ \frac{m^*}{m\varepsilon^2} \tag{13.18}$$

We take the energy of the atom to be in the lowest energy state. Typically, m^*/m is about 0.1 and $\varepsilon \approx 10$. The correction to the band edge energy is only about 0.01 eV.

We learn more using effective mass theory. The orbital radii of the electron–hole pair is changed to the following:

$$a_B = \frac{\varepsilon h^2}{m^* \mathrm{e}^2} = 0.529\text{Å}\ \frac{m^* \varepsilon}{m} \tag{13.19}$$

Roughly speaking, this sets the length scale for an excitation.

For silicon nanocrystals, ε decreases in size from the bulk value. Crudely speaking the dielectric constant decreases as the gap increases. As a consequence of this scaling, small silicon nanocrystals, which have small dielectric constants, possess larger interactions between the electron and hole as the orbital size is reduced. The exciton energy is believed to be on the order of ~1 eV for nanocrystals of less than ~1 nm. So, we cannot compare the quasi-particle gap as computed from the EA and IP directly with experiment and expect to have a quantitative agreement. The quasi-particle gap ignores the electron–hole interaction, which could be large for nanocrystals [5].

13.1.1.3 Vibrational Properties

If we accurately compute the electronic energy as a function of position, we can predict the role of quantum confinement on the vibrational properties of nanocrystals. We expect vibrational properties as a function of size to converge quickly to the bulk values. This reflects the more localized nature of a vibrational mode. Methods such as density functional theory tend to be quite accurate for vibrational modes, which can be computed from small changes in energy with atomic position.

Vibrational modes in nanocrystals can play an important role within the context of the photovoltaic applications. In particular, vibrational properties are directly related to the phonon-assisted optical transitions. This is an important consideration for silicon. In the bulk limit, the lowest optical transition in a silicon crystal is “indirect.” The momentum state of the highest occupied state and the lowest empty state in the crystal are not the same. This situation is peculiar to several elementary semiconducting crystals such as carbon, silicon, and germanium, and compound semiconductors such as gallium phosphide.

Since a photon carries little momentum, the momentum created by the production of an electron–hole pair should also be negligible, should we wish to conserve momentum. In indirect semiconductors such as silicon, the hole is not the issue as it carries almost no momentum. The problem is with the excited electron, which carries momentum. We cannot make an optical excitation in silicon at the indirect gap unless we invoke a new source of momentum to “cancel” out the excess momentum from the electron. Where can we find a low-energy particle that carries momentum? We can find it in “phonons,” or quantized lattice vibrations.

Intuitively, indirect gap materials are not good optical materials as there are four particles involved: a photon, a phonon, an electron, and a hole. The phonon can either be absorbed along with a photon or be emitted with the electron or hole. In any event, to get both momentum and energy conserved means a cooperative effect with an additional particle. The probability of this occurring is notably lower than “direct” transitions.

Owing to the localized nature of nanocrystals, it is feasible to predict vibrational modes calculations by the direct force-constant method. We briefly touched on defining a dynamical matrix for vibrations modes in molecules in Chapter 11. Here we outline the procedure in more detail for nanocrystals. The dynamical matrix of the system is constructed by displacing atoms one by one from their equilibrium positions along the Cartesian directions and finding the forces induced on the other atoms of the nanocrystal. We determine the forces using the Hellmann–Feynman theorem in real space and employ a symmetrized form of the dynamical matrix expression. The elements of the dynamical matrix, $D_{ij}^{\alpha\beta}$, are given by

$$D_{ij}^{\alpha\beta} = -\frac{1}{2}\left[\frac{F_i^{\alpha}\left(\{\vec{R}\} + d_j^{\beta}\right) - F_i^{\alpha}\left(\{\vec{R}\} - d_j^{\beta}\right)}{2d_j^{\beta}} + \frac{F_j^{\beta}\left(\{\vec{R}\} + d_i^{\alpha}\right) - F_j^{\beta}\left(\{\vec{R}\} - d_i^{\alpha}\right)}{2d_i^{\alpha}}\right] \tag{13.20}$$

where F_j^α is the force on atom α in the direction i and $\{\vec{R}\} + d_j^\beta$ is the atomic configuration where the atom β is displaced along j from its equilibrium position. The value of displacement is taken to be small so that we are in the harmonic limit, e.g. it might be 1% of the equilibrium bond length.

The vibrational modes frequencies and corresponding eigenvectors can be obtained from the dynamical equation:

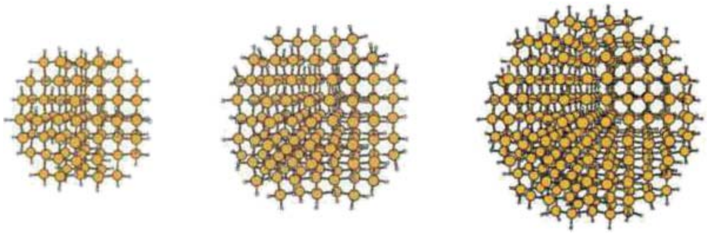
$$\sum_{\beta,k} \left[\omega^2 \delta_{\alpha\beta} \delta_{ik} - \frac{D_{ik}^{\alpha\beta}}{\sqrt{M_\alpha M_\beta}} \right] A_k^\beta = 0 \tag{13.21}$$

where M_α is the mass of the atom labeled by α.

13.1.1.4 Example of Vibrational Modes for Si Nanocrystals

We apply this procedure to several Si nanocrystals: $Si_{87}H_{76}$, $Si_{147}H_{100}$, and $Si_{281}H_{172}$. The interior of these nanocrystals is a bulk-like structure with the surface passivated by hydrogen atoms. The structures are illustrated in Figure 13.4.

In Figure 13.5, we plot the calculated Si vibrational density of states for each nanocrystal. Vibrational density of states contributions from passivating H atoms are excluded as they serve only to terminate the nanocrystal surface in a "gentle" manner. The vibrational density of states of the nanocrystals is compared to the bulk crystal. There are two prominent peaks in the Si vibrational density of states, a low frequency acoustic peak centered in the range of 100–150 cm^{-1} and a higher frequency optical peak in the vicinity of 480 cm^{-1}. "Acoustic" and "optical" refer to the vibrational mode. As one might guess, acoutic modes can



$Si_{87}H_{76}$ (1.4 nm) $Si_{147}H_{100}$ (1.7 nm) $Si_{281}H_{172}$ (2.2 nm)

Figure 13.4 Relaxed structures of silicon nanocrystals used for computing vibrational modes.

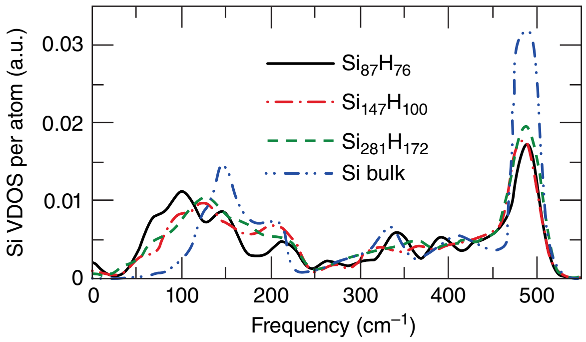


Figure 13.5 The vibrational density of states for nanocrystals in Figure 13.4 and for crystalline silicon.

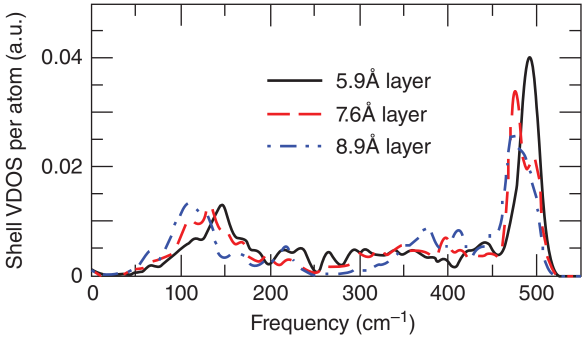


Figure 13.6 Partial vibrational density of states for shells of Si atoms in $Si_{281}H_{172}$ at various distances from nanocrystal center.

propagate sound and optical modes can interact with light. As the nanocrystal decreases in size, the acoustic peak is red-shifted and the amplitude of the high-frequency optical peak is decreased. Both these observations point toward a shift of vibrational density of states to lower frequencies with decreasing nanocrystal size. Why does this happen?

One might suspect the bond lengths in a nanocrystal to be compressed. This often happens when an atom is under-coordinated such as those near the surface of a nanocrystal. Similar bond contractions have also been observed in H-passivated Si (111) surfaces, and these were shown to originate from an electronic charge transfer that tends to reduce the surface dipole moment. Shorter interatomic distances result in stronger Si—Si bonds and larger Si force constants near the nanocrystal surface [6].

Another important effect of confinement on nanocrystal vibrations originates from the effective under-coordination of surface Si atoms. Owing to their small mass, passivating H atoms provide little resistance to the motion of surface Si atoms at frequencies characteristic of Si vibrations, as H atoms follow closely the motion of the Si atoms they are bonded to. This lowering of coordination number leads to a softening of nanocrystal vibrational modes relative to the corresponding modes in bulk.

The total shift in vibrational density of states frequencies is due to the competing influences of surface atom under-coordination and bond length contraction, and their combined effect can be seen by plotting the partial vibrational density of states of shells of Si atoms at various distances from the nanocrystal center, shown in Figure 13.6 for the $Si_{281}H_{172}$ nanocrystal.

Example 13.1 ***Raman Spectra of Silicon Nanostructures*** Raman spectroscopy is a powerful tool for "finger printing" the nature of chemical bonds in materials or molecules. The process involves inelastic scattering of light, usually laser light, which interacts with vibrational modes with the incident light being shifted up or down in energy. In Figure 13.7 the Raman process is illustrated. Raman scattering will result in the scattered light being lower in energy (Stokes Raman scattering) or higher in energy (Anti-Stokes Raman scattering). Rayleigh scattering does not change the photon energy [7].

Raman spectroscopy is based on an interaction between the electron structure of a material and the external electrical field of light, which can create an induced dipole moment within the material based on its polarizability. The details of this

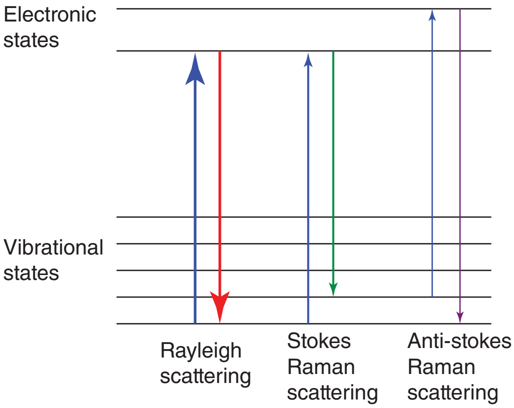


Figure 13.7 Schematic energy levels for Raman spectroscopy. The absorption or emission of vibrational energy distinguishes Stokes from Anti-Stokes scattering.

interaction are complex and beyond the scope of this text. For our purposes, we need only know that the vibrational modes of a nanocrystal are important in determining the energy differences in Raman along with the details of the chemical bond.

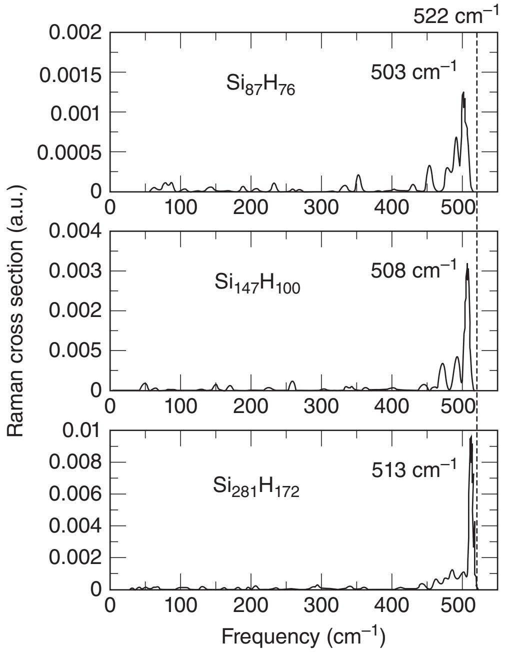
Experimental studies based on Raman spectroscopy have been performed to study the vibrational properties of Si nanocrystals [8]. The main optical Raman peak is found to be increasingly red-shifted and broadened as the nanocrystal size decreases. These models center on confinement effects and a relaxation of the wave-vector selection rule. Selection rules exist when a crystal has translational symmetry. These rules become inoperative in a nanocrystal, which lacks such symmetry. With the ability to compute the electronic and vibrational properties of nanocrystals, it is not necessary to rely on phenomenological theories that often transfer parameters from the bulk crystal to a nanocrystal.

Figure 13.8 illustrates the Raman spectra for the nanocrystals illustrated in Figure 13.4. The major bulk Raman peak is shown at 522 cm^{-1}. This peak red-shifts from the largest nanocrystal ($Si_{281}H_{172}$) at 513 cm^{-1} to the smallest nanocrystal ($Si_{87}H_{76}$) at 503 cm^{-1}. We attribute the shift to the vibrational modes weakening near the surface of the nanocrystal in Figure 13.6. These modes tend to be the dominant ones when one includes the interaction of light with the polarizability modes.

13.1.2 Extrinsic Properties of Silicon Nanocrystals

Doping a small percentage of foreign atoms in bulk semiconductors can profoundly change their electronic properties. Phosphorus-doped crystalline Si introduces defect energy states close to the conduction band [9]. Here "close" means on the order of the thermal energy associated with the room temperatures, i.e. ~0.025 eV. For such "shallow" donors, electrons can be easily thermally excited, thereby greatly enhancing the conductivity of the original pure semiconductor, often by orders of magnitude at room temperature. Doping with boron also introduces defect levels. However, unlike phosphorus, boron results in undersaturated bonds, which create hole states near the top of the valence

Figure 13.8 Calculated Raman cross sections for Si nanocrystals. The dotted line represents the frequency of Raman peak position for bulk-Si.



band. Managing the doping properties of defects plays a vital role in designing modern electronic devices. The design of electronic devices, each smaller than the previous generation, advances the performance of electronic circuits.

As this is written, device miniaturization is approaching the nanometer-scale, if not the sub-nanometer scale. As a consequence, it is important that one understands how doping operates at the nanoscale where confinement is expected to modify the current laws of device design – particularly, at what length scale device design based on macroscopic laws will be altered by quantum theory as noted in the quote by Feynman.

13.1.2.1 Example of Phosphorus-Doped Silicon Nanocrystals

Phosphorus-doped silicon nanocrystals represent the stereotypical system for studying impurities in nanocrystals. Unfortunately, theoretical studies of shallow impurities in nanocrystals are computationally challenging. Only within the last few years have computations been practical for Si nanocrsytals that are comparable in size to experiment.

Suppose we wish to substitute a P atom for a Si atom in a nanocrystal. In bulk Si, the solubility limit is a few atomic percent of P. This suggests that for very small nanocrystals, e.g. a hundred Si atoms, we might start to approach the limit. Of course, there are issues in assuming that the nanocrystal will behave like the bulk phase. Still, if we assume that this limit holds at the nano scale, one might have difficulty doping nanocrystals. Some have postulated that nanocrystals might *self-purify* by expelling dopants under such circumstances [10]. A dopant in a crystal might hop to a different site, but even if the dopant hops hundreds or

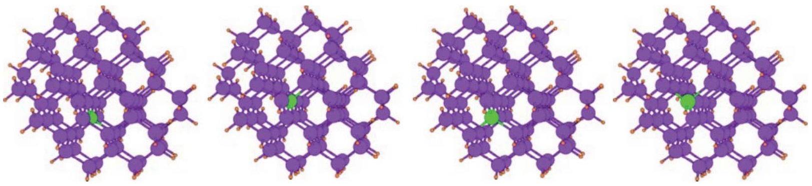


Figure 13.9 Schematic example of the self-purification process in doped Si nanocrystal. By a few diffusion hops the impurity in the interior of the nanocrystal can migrate to the surface, leaving the interior free of any defects.

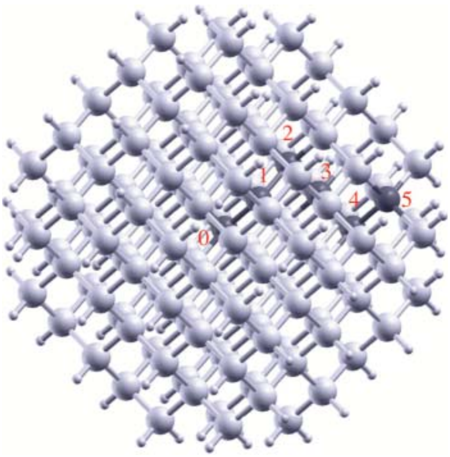


Figure 13.10 A perspective view of a Si nanocrystal. The lightly shaded atoms correspond to Si atoms, and the small atoms on the surface of the nanocrystal are H atoms. The heavily shaded atoms correspond to the Si atoms that will be substituted by P. The numbers measure the distance away from the origin in the unit of Si bond length.

thousands of times it would not reach the surface of a crystal unless it were in the vicinity. In a nanocrystal, this is not the case. If the nanocrystal is small, any atom in the interior might reach the surface in a few hops. This process is illustrated in Figure 13.9.

We illustrate this issue by considering the binding energy of the dopant atom as a function of the P atom in the nanocrystal. In Figure 13.10, we label some atomic sites starting in the center and move out along a bond length to one removed by five bond lengths along the [100] direction.

To be cautious about the nature of the nanocrystal surface, we do not replace Si atoms residing on the surface of the nanocrystal by P. These atoms are passivated by H atoms and we do wish to consider details of the surface structure. For P atoms in the interior, we relax the atoms in the nanocrystal to find the lowest energy structure. Our results for five small nanocrystals of increasing size are shown in Figure 13.11. For Si nanocrystals with a diameter smaller than ~2 nm (the number of Si atoms less than 100), P tends to substitute Si near the surface, e.g. P in $Si_{70}H_{84}$; the energy is lowered monotonically toward the surface of the nanocrystal. Otherwise, the center of the nanocrystal is the energetically most favorable position. This suggests that a “critical size” exists for nanocrystals. Below this size, P atoms will be energetically expelled toward the surface.

Since the substituted P atom creates strain within the Si nanocrystal, expelling the P atom toward the surface will reduce the strain energy. One might also argue

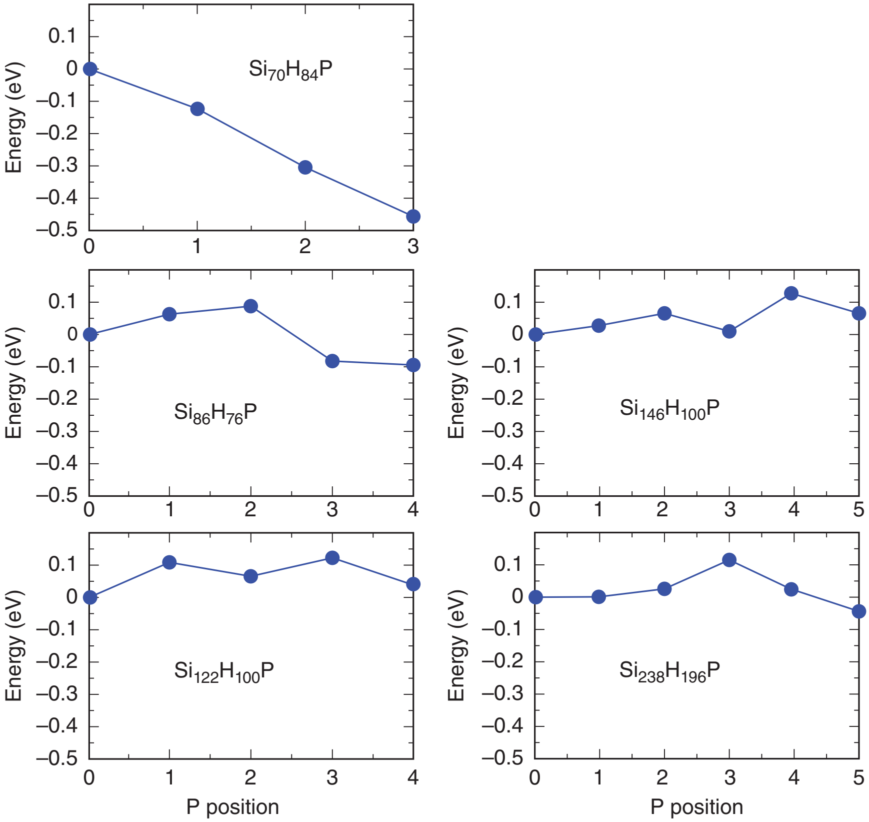


Figure 13.11 Difference in energy(left) and HFS(right) as the P atom moves away from the center of the Si nanocrystal. The energies are with respect to the energy of the Si nanocrystal with P at the center. The *x*-axis measures the distance of P atom away from the origin as illustrated in Figure 13.10.

that if the P atom is not at the center of the nanocrystal, the defect wave function is distorted around P, leading to a loss in Coulomb energy between the P ion and the defect electron. For a “sufficiently large” nanocrystal the P atom will be energetically stable in the interior. Sufficiently large can be defined on the basis of our computations. For nanocrystals containing more than a hundred Si atoms or so, a basin exists at the center of the nanocrystal. This basin spanning several bond lengths provides a local, if not absolute, minimum for the P atom near the center of the nanocrystal. While our focus is on P in Si nanocrystals, these trends appear similar for other dopants such as B in Si.

The position of dopant levels is another difficult problem. In principle, excited states can be computed with some accuracy using approaches beyond density functional theory, albeit with some effort. Still, one can get an idea of how dopant levels change from small nanocrystals to larger ones by computing the energy to excite electrons with density functional theory.

Consider the following: We have a P defect level that resides in the energy gap of a nanocrystal. The ionization energy of the P-Si nanocrystal, I(P-Si nc), represents the energy required to remove a donor electron from this level to the vacuum. Now consider the energy to add an electron to the system for an intrinsic Si nanocrystal. We label this as the EA of the Si nanocrystal: A(Si nc). The difference between I and A yields an estimate of the binding energy of the P electron:

$$E_b = I(\text{P-Si nc}) - A(\text{Si nc}) \tag{13.22}$$

The formula is flawed in simple density functional theory, e.g. there is interaction between the excited electron and the hole. Nonetheless, the trends are interesting.

In Figure 13.12, we plot the ionization energy for the doped nanocrystal and the EA for the intrinsic nanocrystal. The electron affinity of pure Si nanocrystal increases with the size of the nanocrystal. We might guess that adding an electron to a small nanocrystal costs more energy owing to localizing an electrostatic charge. The ionization energy of the P electron depends weakly on nanocrystal size, i.e. it is effectively pinned close to the work function of bulk silicon (~4 eV). Perhaps this is not surprising as the P electron is fairly localized and does not change that much with the size of the nanocrystal.

In Figure 13.13, we illustrate the size dependence of the binding energy E_b of P-doped Si nanocrystal. In small nanocrystals, the defect electron is found to be more tightly bound as nanocrystal size decreases. This is largely due to the weak

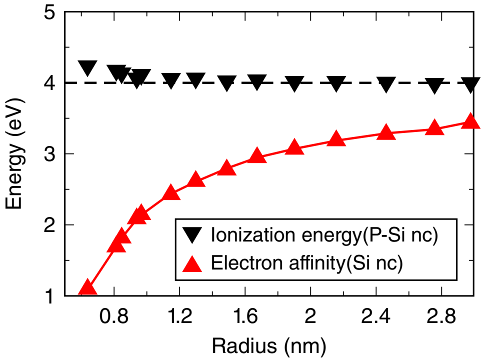


Figure 13.12 Ionization energy of P-doped Si nanocrystal and electron affinity of pure Si nanocrystal plotted as a function of nanocrystal radius.

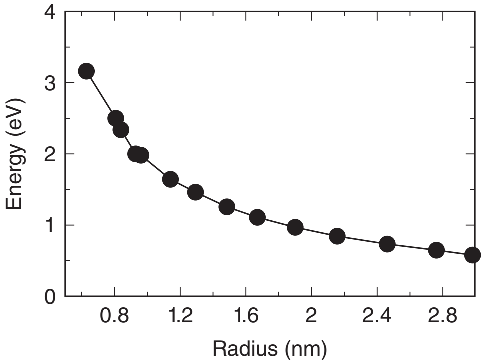


Figure 13.13 The size dependence of the binding energy E_b of P-doped Si nanocrystal.

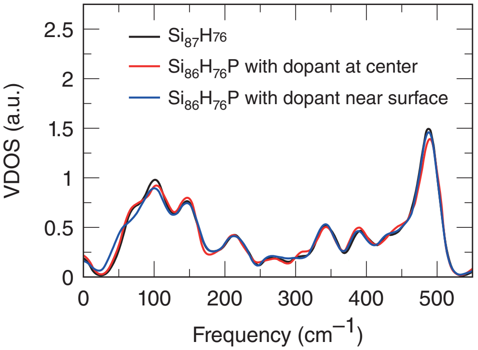


Figure 13.14 The vibrational density of states $Si_{87}H_{76}$ and $Si_{86}H_{76}P$ nanocrystals with P dopant near the surface and at the center positions.

screening in nanocrystals. For nanocrystals up to ~6 nm in diameter, the binding energy remains significantly larger than kT, which implies that the dopant is not "shallow." A fit to the calculated binding energy E_b scales as $R^{-1.1}$. By extrapolating E_b to a large radius, E_B tends to a limiting "bulk" value of ~0.06 eV. Given the simplicity of the computations, this value is pretty close to the experimental value of 0.044 eV for bulk silicon.

Earlier, we demonstrated how Raman spectroscopy can be used to quantify the role of quantum confinement on vibrational modes in silicon nanocrystals. What about the role of dopant atoms? Can we use Raman to analyze dopants, and maybe even determine where dopants reside?

Suppose we replace a Si atom by a P in a silicon nanocrystal. We can compute the dynamical matrix for different sites. In Figure 13.14, we illustrate the results for the $Si_{87}H_{76}$ nanocrystal. At first, the results seem odd as nothing of note happens. Most of the vibrational modes remain unchanged. We can compute the force constants for P and Si atoms in equivalent sites and verify that this is the case. For the center atom of the nanocrystal, the force constant associated with an infinitesimal displacement of a Si atom and a P atom are listed in Table 13.1, along with the corresponding atomic masses.

The force constants are very similar for both atoms. The constant for P is smaller and we expect that. First, the mass of the atom is slightly higher than that of Si. Second, P has an extra electron. There is no place for the extra P electron – all the bonds are saturated. This weakens the bonds between the P atom and neighboring Si atoms. This can indeed be verified on close inspection of the peaks at 100 and 480 cm^{-1} in Figure 13.14, with the curves for doped nanocrystals peaking at slightly lower frequencies than that of the undoped case. One disappointing finding is that if move the P atom from the center

Table 13.1 Force constants and atomic masses of the Si and P atoms at the center of the $Si_{87}H_{76}$ and $Si_{86}H_{76}P$ nanocrystals.

	Force constant (eV Å^{-2})	Atomic Mass (amu)
Si	14.3	28.1
P	12.9	31.0

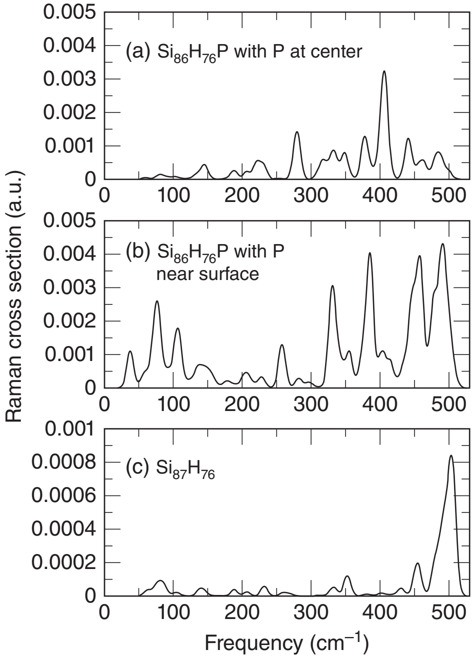


Figure 13.15 Raman cross section for $Si_{87}H_{76}$ and $Si_{86}H_{76}P$ nanoclusters with P dopant at center and near the surface.

of the nanocrystal to the surface, again we do not find significant differences (Figure 13.14). This means that the vibrational modes do not offer a means to either detect or locate the P atom.

This is not the situation for the Raman spectra [11]. In Figure 13.15, the Raman spectra are illustrated with and without a P dopant. Two dopant cases are presented, with the P atom at the center of the nanocrystal and near the surface.

The Raman spectrum for the undoped case is dominated by a single prominent peak at 503 cm^{-1} originating from the so-called Γ-point optical mode. (The Γ-point notation means the vibration mode carries no vibrational momentum.) This frequency is slightly red-shifted relative to the bulk-Si value of 522 cm^{-1} owing to vibration softening and selection rule effects in a finite system.

The introduction of a P dopant leads to qualitative changes in the calculated Raman spectra. The number as well as positions of the peaks are changed depending on the position of the dopant in the nanocrystal. For the center-doped nanocrystal, the main feature in the spectrum is a peak at about 400 cm^{-1}. For the surface-doped case, there are multiple peaks appearing at the frequencies 70, 320, 380, 450, and 480 cm^{-1}.

In Figure 13.14, we established that the vibrational modes are little influenced by the introduction of dopants. The large changes in the Raman spectra must come from changes to the electronic properties. We will not go into the detail of this process. They can be found elsewhere. In brief, the derivative of the electric susceptibility with respect to nuclear coordinates plays an important role in Raman. This assumption is consistent with the increase in magnitude of the Raman spectra, as we expect P doping to increase the responsiveness of Si to external electric fields. What is important is that Raman can be used as a tool to predict where dopants may reside whereas the vibrational modes cannot.

References

1. Toumey, C. (2005). Apostolic succession. *Engineering & Science* 1 (2): 16.
2. Chelikowsky, J.R. (2002). Why silicon is the benchmark. *Materials Today* 5: 64.
3. Sham, L.J. and Schlüter, M. (1983). Density functional theory of the energy gap. *Physical Review Letters* 51: 1888.
4. Tiago, M.L. and Chelikowsky, J.R. (2006). Optical excitations in organic molecules, clusters and defects from first principles Green's function methods. *Physical Review B* 73: 205334.
5. Ogut, S., Chelikowsky, J.R., and Louie, S.G. (1997). Quantum confinement and optical gaps in Si nanocrystals. *Physical Review Letters* 79: 1770.
6. Khoo, K.H., Zayak, A.T., Kwak, H., and Chelikowsky, J.R. (2010). First-principles study of confinement effects on the Raman spectra of Si nanocrystals. *Physical Review Letters* 105: 115504.
7. Gardiner, D.J. (1989). *Practical Raman Spectroscopy*. Springer-Verlag.
8. Richter, H., Wang, Z.P., and Ley, L. (1981). The one phonon Raman spectrum in microcrystalline silicon. *Solid State Communications* 39: 625.
9. Kittel, C. (2005). *Introduction to Solid State Physics*, 8e. Wiley.
10. Chelikowsky, J.R., Chan, T.-L., Alemany, M.M.G., and Dalpian, G. (2011). Computational studies of doped nanostructures. *Reports on Progress in Physics* 74: 046501.
11. Khoo, K.H. and Chelikowsky, J.R. (2014). First-principles study of vibrational modes and Raman spectra in P-doped Si nanocrystals. *Physical Review B* 89: 195309.

Appendix A

Units

Units should be a trivial part of any problem. Also, they should expedite an understanding of the magnitude of the property and ease a comparison with experiment.

We will use *atomic units* (*a.u.*) from time to time and almost exclusively in MATLAB codes. Atomic units are "sensible" ones for systems within the microscopic or atomistic regime.

Consider the following example. Suppose we want to know the Coulombic energy between two electrons that are separated by 1 Å. In the standard SI (or MKS) units, we could calculate this energy from

$$E = -\frac{e^2}{4\pi\epsilon_0 r} \tag{A.1}$$

where e is the charge on the electron in Coulombs (C), ϵ_0 is the permittivity of free space in Farads per meter (F m^{-1}), and r is the separation of the electrons in meters m. We need to know the following:

$$\begin{aligned} e &= 1.60 \times 10^{-19}\ C \\ \epsilon_0 &= 8.85 \times 10^{-12}\ \text{F m}^{-1} \\ 1\text{Å} &= 1 \times 10^{-10}\ \text{m} \end{aligned}$$

The energy in question in Joules (J) is

$$E = -\frac{(1.60 \times 10^{-19})^2}{4 \times 3.142 \times 8.85 \times 10^{-12} \times 10^{-10}} \quad \text{J} = 2.30 \times 10^{-18} \quad \text{J} \tag{A.2}$$

This is a perfectly acceptable result, but it can be hard to make comparisons. We could try to make a comparison with a macroscopic situation. We can consider the change in potential energy for a mass of 1 kg dropped from a height of 1 m. This energy is given as $E = mgh = 1$ kg $\times$ 9.8 m s^2 $\times$ 1 m $=$ 9.8 J. The potential energy for this change differs for two electrons separated by 1 Å by almost 20 orders of magnitude. Such a difference is not within the normal human experience. All we really know from this comparison is that the energy of the two electrons is extremely small compared to the potential energy change of the 1 kg mass. If we are interested in getting a better physical understanding of the electronic energy, we should not use units appropriate for macroscopic objects.

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

Better suited units involve some "standard" at the atomic scale. Suppose we consider a set of units appropriate for atoms. We can use atomic units where we take the following:

The charge on the electron = $e = 1$

Planck's constant divided by $2\pi = \hbar = 1$

Mass of the electron = $m = 1$

In this set of units, energy is measured in Hartrees (E_h), which is 4.36×10^{-18} J, and distance is measured in Bohr units (a_0), which is 0.5292 Å or 5.292×10^{-11} m.

We will also use CGS formulae as the starting point for any electrostatic or magnetic interaction energies. This choice avoids considering the permittivity of free space for electrostatic interactions or the permeability of free space for magnetic interactions.

For the problem at hand, we can write the Coulombic interaction without the permittivity factor and we need only convert 1 Å to 1.890 a_0. We write

$$E = -\frac{e^2}{r} = -\frac{1}{r} = -\frac{1}{1.890} = -0.5292E_h \tag{A.3}$$

This number is the result of a *much simpler* calculation and has a simpler interpretation. Atomic units are based on the hydrogen atom. In Chapter 2, we showed that the potential energy of the hydrogen atom is 1 a.u. in the ground state. This gives us our reference. For the example at hand, the electronic potential energy is 0.529 times that of the potential energy for the electron-potential interaction within a hydrogen atom.

Below we summarize some useful values and conversion factors (Tables A.1 and A.2).

Table A.1 Table of values.

Quantity	SI units	CGS units	Atomic units
Charge on the electron (e)	1.602×10^{-19} C	4.803×10^{-10}	1
Electron rest mass (m)	9.110×10^{-31} kg	9.110×10^{-28} g	1
Planck's constant ($\hbar$)	1.055×10^{-34} J s	1.055×10^{-27} erg s	1
Speed of light (c)	2.998×10^{8} m	2.998×10^{10} cm	137.04
Rydberg constant ($me^4/2\hbar^2$)	2.180×10^{-18} J	2.180×10^{-11} ergs	0.5
Bohr radius ($\hbar^2/me^2$)	5.292×10^{-11} m	5.292×10^{-9} cm	1
1 electronvolt (eV)	1.602×10^{-19} J	1.602×10^{-12} erg	0.03675
Boltzmann constant (k_B)	1.381×10^{-23} JK^{-1}	1.381×10^{-16} erg K^{-1}	3.167×10^{-6} K^{-1}

Table A.2 Conservation factors.

Quantity	Equivalent
1 Hartree (E_h)	27.212 eV
1 kcal mol^{-1}	0.04336 eV
1 K	8.616975×10^{-5} eV
1 cm^{-1}	1.239×10^{-4} eV
1 kJ mol^{-1}	0.01036 eV
1 eV	8,065 cm^{-1}
1 eV	96.49 kJ mol^{-1}
1 eV	23.06 kcal mol^{-1}

Appendix B

A Working Electronic Structure Code

Many of the computations illustrated in this text were not done with an elaborate electronic structure code. Such codes can run in thousands of lines, if not tens of thousands of lines. The codes are not easy for a novice to learn. Instead, many of the computations were run with an abbreviated MATLAB code. MATLAB is a nice coding language. One can write sophisticated codes to address complex problems. Also, MATLAB commands are easy to learn and the user often gets "tips" when running the code as to where a mistake might have been made.

To facilitate the use of this approach, we removed many of the bells and whistles of a sophisticated research code to offer the reader a code that allows one to do reasonable computations, and yet, retain a decent level of accuracy.

The general setup is outlined in Chapter 9. The Kohn–Sham equations are solved self-consistently on a real-space grid within a specified domain. The code is fashioned to avoid having the user set a number of internal parameters or construct pseudopotentials, i.e. the code is preloaded with pseudopotentials for the first two rows of the periodic table.

The MATLAB code and details on how to run it are available on github:

https://github.com/jchelikowsky/Matlab_Real_Space

Example B.1 ***The*** C_{60} ***Molecule*** Richard Smalley, Robert Curl, and Harold Kroto won the 1996 Nobel Prize in Chemistry for the discovery of a new form of carbon. The molecule reminds one of a soccer ball with combinations of five- and six-fold rings. The simplicity of the molecule is striking; yet, this object remained undiscovered until the latter half of the twentieth century. One wonders what other molecules might exist in nature that remain unknown.

At the time of the discovery in 1985, computing the electronic structure of this molecule was a challenge. Solving the Kohn–Sham equations could be done with some effort. In contrast, the problem is not difficult to solve using new algorithms available to us as outlined in our text. The required computational platform is a laptop.

Introductory Quantum Mechanics with MATLAB®: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.

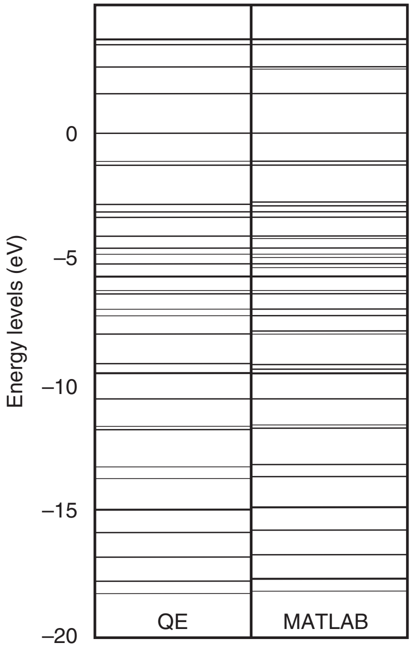
Table B.1 Coordinates for the C_{60} molecule in atomic units.

Atom	*x*	*y*	*z*	Atom	*x*	*y*	*z*
1	0.00	−1.36	−6.81	31	1.33	6.60	−0.04
2	0.00	1.36	−6.81	32	−1.33	6.60	−0.04
3	−2.25	−2.74	−5.90	33	6.60	0.00	1.29
4	2.25	−2.73	−5.89	34	−4.93	4.38	1.34
5	−2.25	2.73	−5.89	35	4.93	4.38	1.34
6	2.25	2.73	−5.89	36	−4.93	−4.38	1.34
7	4.39	1.38	−5.00	37	4.93	−4.38	1.34
8	−4.39	1.38	−5.00	38	2.70	5.76	2.20
9	−4.39	−1.38	−5.00	39	−2.70	5.76	2.20
10	4.39	−1.38	−5.00	40	−2.70	−5.76	2.20
11	−1.38	−4.94	−4.45	41	2.70	−5.76	2.20
12	1.38	−4.93	−4.45	42	−5.76	−2.23	2.68
13	−1.38	4.93	−4.45	43	5.76	−2.23	2.68
14	1.38	4.93	−4.45	44	5.76	2.23	2.68
15	5.74	2.22	−2.75	45	−5.76	2.23	2.68
16	−5.74	2.22	−2.75	46	−1.38	−4.94	4.36
17	−5.74	−2.22	−2.75	47	1.38	−4.94	4.36
18	5.74	−2.22	−2.75	48	−1.38	4.94	4.36
19	−2.70	5.74	−2.27	49	1.38	4.94	4.34
20	2.70	5.74	−2.27	50	4.38	1.38	4.92
21	−2.70	−5.74	−2.27	51	−4.38	1.38	4.92
22	2.70	−5.74	−2.27	52	−4.38	−1.38	4.92
23	−4.92	−4.38	−1.41	53	4.38	−1.38	4.92
24	4.92	−4.38	−1.41	54	−2.23	−2.71	5.74
25	−4.92	4.38	−1.41	55	2.23	−2.71	5.74
26	4.92	4.38	−1.41	56	−2.23	2.71	5.74
27	6.58	0.00	−1.36	57	2.23	2.71	5.74
28	−6.58	0.00	−1.36	58	0.00	−1.34	6.60
29	−1.33	−6.60	−0.04	59	0.00	1.34	6.60
30	1.33	−6.60	−0.04	60	−6.60	0.00	1.29

The first step is to input the structure of the C_{60} molecule for each carbon atom. The atomic coordinates are given in Table B.1 in Bohr units (1 a.u. =0.5292 Å). The MATLAB code will automatically extract the correct pseudopotential at each site, set up the grid based on the atom present, and set up the domain.

The code utilizes filtering to solve the eigenvalue problem using the techniques outlined in Chapter 9. Some of the key references involved in the construction of

Figure B.1 Energy levels of C_{60}. The left panel shows results from the Quantum ESSPRESSO code, and the right panel from our MATLAB code.



this code are listed in Refs. [1–4]. After the initial solution, new wave functions are created. The resulting charge density is used to create new potentials and the process iterated until self-consistent. At this point, we quit and declare a solution. Our goal is to extract a "reasonable" description of the spatial and energetic distribution of the electronic states without the use of a full research code.

How do we know our simple code works? We assess the matter by using a different code with a different approach and comparing the two results. We do not expect perfect agreement as the MATLAB code runs quickly on a laptop versus a full-blown research code. Nonetheless, our results compare favorably in Figure B.1 to a state of the art code called "Quantum ESPRESSO" [5]. The energy gap and band width agree to within ∼ 0.1 eV over a 25 eV span. The MATLAB code is not guaranteed to yield correct symmetry until fully self-consistent. Thus, there are some small shifts for degenerate states.

Once the Kohn–Sham wave functions are known, we plot the charge density of the C_{60} molecule. The charge density is a three-dimensional function. The transcription to a two-dimensional plot can be difficult. In Figure B.2, we plot the charge density as a surface of constant density, mostly for show. However, one can distinguish the difference between single and double bonds.

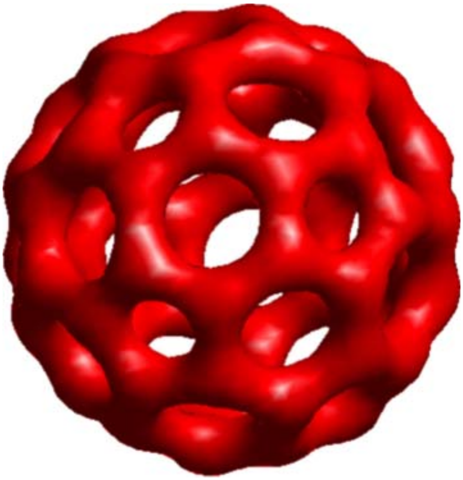


Figure B.2 Isosurface plot of the C_{60} molecular charge density.

References

1 Chelkowsky, J.R., Troullier, N., and Saad, Y. (1994). The finite-difference-pseudopotential method: electronic structure calculations without a basis. *Physical Review Letters* 72: 1240.

2 Kronik, L., Tiago, M.L., Alemany, M.M.G. et al. (2006). PARSEC-the pseudopotential algorithm for real space electronic structure calculations: recent advances and novel applications to nano-structures. *Physica Status Solidi B* 243: 1063.

3 Zhou, Y., Saad, Y., Tiago, M.L., and Chelikowsky, J.R. (2006). Parallel self-consistent-field calculations via Chebyshev-filtered subspace acceleration. *Physical Review E* 74: 066704.

4 Schofield, G., Chelikowsky, J.R., and Saad, Y. (2012). A spectrum slicing method for the Kohn-Sham problem. *Computer Physics Communications* 183: 497.

5 Giannozzi, P., Baroni, S., Bonini, N. et al. (2009). QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter* 21: 395502.

Index

Introductory Quantum Mechanics with MATLAB$^{®}$: For Atoms, Molecules, Clusters, and Nanocrystals,
First Edition. James R. Chelikowsky.